

**DIE LEHREFFEKTIVITÄT VON PROGRAMMIERTER
UNTERWEISUNG UND HERKÖMMLICHEM LEHRERUNTER-
RICHT UNTER BERÜCKSICHTIGUNG ADRESSATEN-
BEZOGENER MASSKRITERIEN**

von

Hellmuth Walter

aus

Tübingen

1971 № 489

**DIE LEHREFFEKTIVITÄT VON PROGRAMMIERTER
UNTERWEISUNG UND HERKÖMMLICHEM LEHRERUNTER-
RICHT UNTER BERÜCKSICHTIGUNG ADRESSATEN-
BEZOGENER MASSKRITERIEN**

Inaugural-Dissertation
der Philosophischen Fakultät
der Friedrich-Alexander-Universität
Erlangen-Nürnberg

vorgelegt

von

Hellmuth Walter

aus

Tübingen



Tag der mündlichen Prüfung: 16. 7. 1971

Dekan: Universitätsprofessor Dr. U. Fülleborn
Referent: Universitätsprofessor Dr. W. Toman
Korreferent: Universitätsprofessor Dr. J. Franke

INHALTSVERZEICHNIS

Digitized by the Internet Archive
in 2026

Einleitung	
1.00 Fragestellung der Untersuchung	1
2.00 Forschungsstand und Forschungsergebnisse aus dem Bereich der Programmierten Unterweisung - eine allgemeine Übersicht	4
2.10 Vergleichsstudien	8
2.20 Experimentelle Analyse der Variablen, welche den Lehreffekt der Programmierten Unterweisung beeinflussen	21
2.21 Untersuchungen, welche den Darbietungsmodus betreffen	23
2.22 Untersuchungen, welche das Darbietungsmittel betreffen	29
2.23 Untersuchungen, welche den Antwortmodus betreffen	31
2.24 Untersuchungen, welche die Wirkungsweise der Verstärkung betreffen	35
2.30 Analyse des Zusammenhangs von Lernerfolg bei Programmierter Unterweisung und Persönlichkeitsvariablen	40
2.40 Zusammenfassung	47
3.00 Bericht und methodenkritische Diskussion repräsentativer Vergleichsstudien	54
3.10 G. GLEASON: "Programed learning in sixth grade arithmetic"	55
3.20 D.M. MILLER, D.E. WILIEY, P. LAMBERT: "An experimental study of the effect of instructional method on the acquisition, retention and transfer of English punctuation skills"	57
3.30 A. ROE, M. MASSEY, G. WELTMAN und D. LEEDS: "Automated Teaching Methods using linear Programs"	59

	Seite
3.40 G. HUBER: "Lernpsychologische Befunde bei Programmierter Unterweisung"	61
3.50 K. RÖTTEL: "Vergleichende Analyse einer Versuchsreihe mit programmiertem und konventionellem Unterricht"	63
3.60 L.J. ISSING und C. SCHELLENBERG: "Anwendung von PU-Prinzipien auf die Gestaltung von Schulfernsehsendungen - eine Vergleichsuntersuchung"	65
3.70 Diskussion der berichteten Untersuchungen	67
3.71 Diskussion der Versuchspläne	67
3.72 Definition der unabhängigen Variable "herkömmlicher Unterricht"	74
3.73 Definition der abhängigen Variablen	77
3.74 Statistische Verrechnungstechniken	81
4.00 Beschreibung und Analyse der Programmierten Unterweisung und des herkömmlichen Unterrichts	83
4.10 Lernpsychologische Grundlagen der Programmierten Unterweisung	83
4.11 Das Prinzip der Verstärkung	85
4.12 Das Prinzip der ausdrücklichen aktiven Reaktion (overt response)	93
4.13 Das Prinzip des "shaping of behavior"	95
4.14 Die Rolle des "einsichtigen Lernens" im System von SKINNER	97
4.20 Begriffsklärung und Analyse des linearen Programmtypus	99
4.30 Allgemeine Erörterung des herkömmlichen Unterrichts	109
4.31 Das frontalunterrichtliche Verfahren	115

	Seite
5.00 Vergleichende Betrachtung der Programmierten Unterweisung und des herkömmlichen Unterrichts	123
5.10 Das Prinzip der Verstärkung	123
5.20 Das Prinzip der "aktiven Antwortformulierung"	138
5.30 Das Prinzip des "shaping of behavior"	144
5.40 Das Prinzip der sozialen Entlastung der Lernprozesse	151
5.50 Das Prinzip des individuellen Lerntempos	153
6.00 Hypothesen	156
6.10 Hypothesen, welche die Ausprägung der Lernleistung betreffen	157
6.20 Hypothesen, welche den Zusammenhang von Intelligenz und Lernleistung betreffen	158
6.30 Positive Verstärkungen	159
6.40 "Unterrichtsbezogene Schüleraktivität"	160
6.50 "Konzentrativer Arbeitshaltung"	161
6.60 "Einstellung zur Schule"	162
6.70 "Einstellung zum Lehrer"	162
7.00 Methode der Untersuchung	163
7.10 Grundsätzliche Probleme der experimentellen Forschung im pädagogischen Bereich	163
7.20 Störvariablen, welche die interne und externe Validität beeinträchtigen	172

	Seite
7.21 Faktoren, welche die innere Gültigkeit beeinträchtigen	173
7.22 Faktoren, welche die äußere Gültigkeit beeinträchtigen	183
7.23 Zusammenfassende Schlußfolgerungen für Methodenvergleiche im Rahmen der Schule	186
7.30 Versuchsplan	188
7.31 Allgemeine Erörterung verschiedener Versuchspläne	188
7.32 Versuchsplan	190
7.40 Versuchsablauf	201
7.41 Vorexperimentelle Phase	201
7.42 Experimentelle Phase	202
7.43 Instruktionen	203
7.50 Untersuchte Stichprobe	204
7.60 Verwendete Unterrichtsprogramme	209
7.70 Der "herkömmliche Vergleichsunterricht"	211
7.71 Die Mikrostruktur unseres "herkömmlichen Vergleichsunterrichts"	213
7.72 Protokollierung des Unterrichtsablaufs	216
7.73 Zuverlässigkeit der Erfassung und Validität der Beobachtungskategorien	222
7.80 Maßkriterien der Untersuchung	230
7.81 Unterrichtsbezogene Maßkriterien	230
7.811 Grundsätzliche Bemerkungen	230
7.812 Test zur Ermittlung der Vorkenntnisse	234
7.813 Das Problem der inhaltlichen Formulierung der Testitems	237
7.814 Gewichtung der items	240
7.815 Herstellung der Testendform	242
7.816 Testvorgabe	246

	Seite
7.82 Adressatenbezogene Maßkriterien	258
7.821 Intelligenz	258
7.822 "Einstellung zur Schule"	258
7.823 "Einstellung zum Lehrer"	259
7.824 "Konzentrativer Arbeitshaltung"	259
7.825 Positive Verstärkungen und "stoffbezogene Schüleraktivität"	262
7.90 Abschließende Diskussion methodischer Probleme	262
8.00 Versuchsergebnisse	269
8.10 Hypothesen, welche die Ausprägung der Lernleistung betreffen	269
8.11 Normalität der Verteilungen	269
8.12 Varianzen der Lernleistung	272
8.13 Lernerfolg in der Lernzielkategorie "knowledge"	274
8.14 Lernerfolg in der Lernzielkategorie "comprehension"	277
8.20 Hypothesen, welche den Zusammenhang von Intelligenz und Lernleistung betreffen	279
8.21 Normalität der Verteilungen	279
8.22 Intelligenz und Lernerfolg in der Lernzielkategorie "knowledge"	280
8.23 Intelligenz und Lernerfolg in der Lernzielkategorie "comprehension"	286
8.30 Hypothesen, welche die "positiven Verstärkungen" betreffen	288
8.31 Normalität der Verteilungen	288
8.32 Das Ausmaß an positiven Verstärkungen im herkömmlichen Unterricht und die Lernleistung in den Lernzielkategorien "knowledge" und "comprehension"	288

	Seite
1.40 Hypothesen, welche die "unterrichtsbezogene Schüleraktivität" betreffen	290
1.41 Normalität der Verteilungen	290
1.42 Das Ausmaß an "unterrichtsbezogener Schüleraktivität" im herkömmlichen Unterricht und die Lernleistung in den Lernzielkategorien "knowledge" und "comprehension"	291
1.50 Hypothesen, welche die "konzentrativen Arbeitshaltung" betreffen	293
1.51 Normalität der Verteilungen	293
1.52 "Konzentrativen Arbeitshaltung" und Lernleistung in den Lernzielkategorien "knowledge" und "comprehension"	293
1.60 "Einstellung zur Schule"	296
1.61 Normalität der Verteilungen	296
1.62 Die "Einstellung zur Schule" und die Lernleistung in den beiden Lernzielkategorien	296
1.70 Einstellung zum Lehrer	298
1.71 Normalität der Verteilungen	298
1.72 Die Einstellung zum Lehrer und die Lernleistung in den beiden Lernzielkategorien	298
1.00 Diskussion der Versuchsergebnisse	300
1.10 Genereller Lern Leistungsvergleich	301
1.11 Lernzielkategorie "knowledge" (H. 1-4)	301
1.12 Lernzielkategorie "comprehension" (H 5-8)	302
1.20 Zusammenhang von Intelligenz und Lernleistung	304
1.21 Lernzielkategorie "knowledge" (H 9-11)	304
1.22 Lernzielkategorie "comprehension" (H 12-14)	306

	Seite
9.30 Zusammenhang von Lernleistung und der Anzahl "positiver Verstärkungen" im herkömmlichen Unterricht (H 15-16)	307
9.40 Zusammenhang von Lernleistung und "unterrichtsbezogener Schüleraktivität" (H 17-18)	311
9.50 Der Zusammenhang von "konzentrativer Arbeitshaltung und der Lernleistung in den beiden Unterrichtsmethoden (H 19-20)	312
9.60 Der Zusammenhang zwischen der Einstellung der Schüler zu Lehrer und Schule und der Lernleistung (H 21-22)	314
9.70 Zusammenfassung	318
Literatur	323
Anhang	343

Einleitung

Die vergleichende Betrachtung von Programmierter Unterweisung und herkömmlichem Unterricht hat in dem vergangenen Jahrzehnt im Rahmen der experimentellen Unterrichtsforschung einen zentralen Platz eingenommen. Die Bedeutung der einschlägigen Forschungsbemühungen dürfte sich insbesondere im deutschsprachigen Raum angesichts der geplanten Reformen des Bildungssystems noch vergrößern, da in Zukunft in weit größerem Umfang, als dies gegenwärtig der Fall ist, der Einsatz von programmiertem Unterrichtsmaterial notwendig sein wird. Dabei wird für definierte Programmtypen, bestimmte Adressatengruppen und unterrichtete Stoffe zu prüfen sein, inwieweit der Einsatz programmierten Arbeitsmaterials gegenüber herkömmlichen Unterrichtsverfahren eine vergleichbare bzw. höhere Lehreffektivität erwarten läßt, und inwieweit der herkömmliche Unterricht selbst durch Einbezug und Anwendung von Funktionsprinzipien der Programmierten Unterweisung bezüglich seiner Lehreffektivität verbessert werden kann. Schließlich wird man sich - angesichts der geplanten Individualisierung der Lernsituationen - fragen müssen, bei welchen Schülergruppen der Einsatz programmierten Lernmaterials besonders wirksam ist. Es muß schon an dieser Stelle darauf hingewiesen werden, daß die Klärung all dieser Fragen nicht durch einige grundlegende Arbeiten zu leisten ist, da die Frage der Lehreffektivität einer Unterrichtsmethode immer nur im Hinblick auf ein bestimmtes Stoffgebiet, eine definierte Adressatengruppe und einen bestimmten Programmtypus zu entscheiden ist. Unsere Arbeit ordnet sich im Rahmen der oben genannten Forschungsbemühungen ein. Dabei waren wir auf die Mitarbeit der an der Untersuchung beteiligten Lehrer und Studenten angewiesen, für deren Unterstützung wir uns an dieser Stelle herzlich bedanken.

1.00 Fragestellung der Untersuchung

Eine kritische Analyse der Untersuchungen, welche sich mit der vergleichenden Betrachtung von Programmierter Unterweisung und herkömmlichem Unterricht befassen, zeigt, daß diese fast durchgängig in dreifacher Hinsicht unbefriedigend sind. Einmal fällt auf, daß die (zunächst hypothetisch gesetzte) Überlegenheit der Programmierten Unterweisung ziemlich unkritisch insbesondere auf die von SKINNER und seinen Anhängern formulierten Funktionsprinzipien zurückgeführt wird, obwohl in der Zwischenzeit genügend Hinweise dafür vorliegen, daß die Anwendung der Modelle der amerikanischen Lerntheoretiker auf die meist komplexere Wirklichkeit schulischer Lernsituationen ziemlich problematisch ist und keinesfalls unkritisch vollzogen werden kann. So werden insbesondere die "regelmäßige und sofortige Verstärkung" und die "aktive Antwortformulierung" herangezogen, um die Überlegenheit der Programmierten Unterweisung zu begründen, obwohl gerade diesbezüglich im amerikanischen Bereich Arbeiten mit sehr widersprüchlichen Ergebnissen vorliegen. Diese unkritische Übernahme von Behauptungen, welche bislang i.allg. nur in relativ einfach strukturierten Lernsituationen, verifiziert werden konnten, ist umso bedenklicher, als gleichzeitig darauf verzichtet wird, den experimentell gesetzten herkömmlichen Vergleichsunterricht zu definieren und zu analysieren. Die Frage, inwieweit der herkömmliche Vergleichsunterricht die zur Begründung der Überlegenheit der Programmierten Unterweisung bemühten Prinzipien erfüllt (und darüber hinaus möglicherweise sogar bessere Voraussetzungen bietet), wird seltsamerweise nie gestellt. Zum dritten fällt schließlich auf, daß trotz

der ausgeprägten methodischen Schwierigkeiten von feld-experimentellen Untersuchungen dieselben, von einigen in jüngster Zeit erschienenen methodenkritischen Beiträgen abgesehen, kaum diskutiert werden. Die einschlägigen Untersuchungen sind fast durchgehend in Form einer einfachen Kontrollgruppenanordnung angelegt, welche oben-drein - wenn auch gezwungenermaßen - nicht auf einer Zufallsstichprobe sondern auf geschlossenen Schulklassen basiert. Angesichts dieser grundsätzlichen Mängel der bislang veröffentlichten Vergleichsuntersuchungen ist es umso erstaunlicher, daß die - in erschreckendem Maß widersprüchlichen - Ergebnisse derselben oft in mehr als großzügiger Weise interpretiert und generalisiert werden.

- Ziel der hier vorgetragenen Vergleichsuntersuchung ist es,
- die bislang unkritisch vorgetragenen Begründungen der Überlegenheit der Programmierten Unterweisung zu überprüfen;
 - den experimentell gesetzten herkömmlichen Vergleichsunterricht sauber zu definieren und zu analysieren, um aus einer solchermaßen fundierten vergleichenden Betrachtung der experimentell gesetzten Unterrichtsmethoden unsere Hypothesen abzuleiten;
 - das bisherige Vorgehen der Autoren von Vergleichsstudien methodenkritisch zu analysieren und ein neues Untersuchungskonzept vorzuschlagen, welches zwar, wie alle feldexperimentell angelegten Untersuchungen, welche mit natürlichen Klasseneinheiten arbeiten, Mängel aufweist, gegenüber den bisherigen Verfahren jedoch eine deutliche Verbesserung darstellt.

Auf diesem Hintergrund hoffen wir zu Ergebnissen zu gelangen, welche zuverlässiger interpretierbar sind, als dies bei den bisherigen Vergleichsstudien der Fall ist.

Dabei beschränken wir uns nicht auf den generellen Vergleich der Lehreffektivität unserer beiden Unterrichtsmethoden, sondern versuchen darüber hinaus die Lehreffektivität bei definierten Adressatengruppen vergleichend zu analysieren. Dies scheint uns angesichts der nur ungenügend überprüften Behauptung, daß die Programmierte Unterweisung für bestimmte Adressatengruppen besonders lehreffizient sei, unbedingt notwendig zu sein, zumal in der Schulpraxis entsprechende Gruppendifferenzierungen bereits vorgenommen werden.

Die inhaltliche Gestaltung und Formulierung unserer Ausführungen haben wir so gewählt, daß auch wissenschaftlich interessierte Schulpraktiker, welche keine - insbesondere methodische - Ausbildung im Sinne experimentell orientierter Sozialwissenschaften erhalten haben, sich ohne größere Schwierigkeiten orientieren können. Die z.T. vorhandene Redundanz unserer Ausführungen ist also bewußt intendiert. Dies scheint uns angesichts der Tatsache, daß die zukünftige Unterrichtsforschung in verstärktem Maß auf die Mitarbeit von Schulpraktikern angewiesen sein wird, unerläßlich zu sein.

2.00 Forschungsstand und Forschungsergebnisse aus dem Bereich der Programmierten Unterweisung - eine allgemeine Übersicht

Als S.L. PRESSEY 1926¹⁾ und 1927²⁾ mit ersten Berichten über Lehrmaschinen an die Öffentlichkeit trat, fand er wenig Resonanz. In seinem Aufsatz "A simple apparatus which gives tests and scores and teachers" beschrieb er die Funktion einer einfachen Maschine zur automatischen Intelligenzprüfung, welche Testaufgaben und damit in weiterem Sinne Information darbot und gleichzeitig die Reaktion des Probanden auswertete. Schon in dieser Arbeit wies PRESSEY darauf hin, "daß die Lehrer zur Zeit mit Routine und Schreibarbeiten überlastet sind, die sich auch maschinell erledigen lassen - und dadurch würde der Lehrer frei werden für eigentliches Lehren, für das Vermitteln von Denkanstößen und Idealen"³⁾. In seiner Veröffentlichung von 1927 sagte er eine "industrielle Revolution in der Erziehung" voraus; bereits in den dreißiger Jahren stellte er jedoch resignierend fest, daß in Fachkreisen und in der Öffentlichkeit seine Gedanken nicht aufgegriffen und weitergeführt wurden. "Presseys Maschinen unterlagen zum Teil der kulturellen

-
- 1) S.L. PRESSEY: "A simple apparatus which gives tests and scores and teachers", in: W. CORRELL (Hrsg.): "Programmiertes Lernen und Lehrmaschinen", Braunschweig 1966
 - 2) S.L. PRESSEY: "A machine for automatic teaching of drill material", in: W. CORRELL (Hrsg.): "Programmiertes Lernen und Lehrmaschinen", Braunschweig 1966
 - 3) S.L. PRESSEY: "A simple apparatus which gives tests and scores and teachers", in: W. CORRELL (Hrsg.): "Programmiertes Lernen und Lehrmaschinen", Braunschweig 1966, S. 31

Trägheit. Die Welt der Pädagogik war noch nicht reif für sie. Aber sie hatten auch Schwächen, die vermutlich zu ihrem Mißerfolg beitrugen. Pressey arbeitete auf dem Boden einer psychologischen Theorie, die den Lernprozeß nicht in den Griff bekommen hatte ... die Lernrate wurde zwar beobachtet, aber es wurde wenig dafür getan, sie zu verändern. Warum die Versuchsperson eines solchen Experiments gleichzeitig lernte, war von geringem Interesse¹⁾. Erst der bahnbrechende Artikel von B.F. SKINNER "The science of learning and the art of teaching" war Ausgangspunkt für eine heute kaum mehr übersehbare Anzahl von Arbeiten über die Programmierete Unterweisung, welche allein in den Vereinigten Staaten schon in den sechziger Jahren die Zahl tausend bei weitem überschritt und welche in immer größerem Ausmaß auch in anderen Ländern erscheinen²⁾. Heute kann die Programmierete Unterweisung als das am intensivsten erforschte Gebiet der experimentellen Pädagogik bezeichnet werden. Voraussetzung für diesen Erfolg waren in erster Linie die Lerntheorie von B.F. SKINNER und später - insbesondere in Europa - das durch die angewandte Kybernetik geweckte Verständnis für Regelungsfragen im Bereich der Pädagogik³⁾. Im Gegensatz zu

-
- 1) B.F. SKINNER: "Lehrmaschinen", in: W. CORRELL (Hrsg.): "Programmiertes Lernen und Lehrmaschinen", Braunschweig 1966, S. 39
 - 2) In diesem Zusammenhang sei insbesondere auf die Sowjetunion verwiesen. Entsprechende Hinweise finden sich in der Veröffentlichung von H. VOIGT: "Pädagogische Kybernetik und Programmierte Instruktion in der Sowjetunion und in der DDR 1965/66 mit Bibliographie", Weinheim 1968
 - 3) Repräsentativ für die einschlägigen Bemühungen im deutschsprachigen Raum sind die Veröffentlichungen von F. v. CUBE: "Kybernetische Grundlagen des Lernens und Lehrens", Stuttgart 1968 sowie H. FRANK: "Kybernetische Grundlagen der Pädagogik", Baden-Baden 1962

PRESSEY gelang es SKINNER, an seiner in sich stringenten und bester behavioristischer Tradition entstammenden Lerntheorie anknüpfend, einen unmittelbaren Bezug zur unterrichtlichen Praxis und - in weiterem Sinn - zu "alltäglichen" Lernprozessen überhaupt herzustellen. Vor allem ab 1960 wurde deutlich, daß die unmittelbar auf der SKINNERschen Lerntheorie aufgebaute Programmier- te Unterweisung im Gegensatz zu den herkömmlichen Unterrichtsverfahren den Prinzipien der kybernetischen Pädagogik weitgehend entspricht. Weiter wurde die Forschungsarbeit auf dem Gebiet der Programmierten Unterweisung durch allgemeine Trends in der Erziehungstechnologie und der Schulreform begünstigt, wobei hier allerdings - insbesondere was die Unterrichtstechnologie anbetrifft - von einem Wechselwirkungsverhältnis gesprochen werden muß. Der enge Bezug von Schulorganisation, Unterrichtstechnologie und Programmierter Unterweisung wird u.a. aus der Arbeit von W. ZIFREUND¹⁾ deutlich, aber auch die Vielzahl der Veröffentlichungen über Lehrmaschinen weisen darauf hin. Die Unterrichtstechnologie versucht schon seit längerer Zeit, pädagogische Maßnahmen rationaler als bislang anzuordnen, das Schülerverhalten systematischer zu steuern und das Lerntempo sowie den Schwierigkeitsgrad der Lernstoffe den einzelnen Schülern bzw. Schülergruppen anzugleichen²⁾. Die Programmier- te Unterweisung kommt diesen Trends in jeder Hinsicht entgegen. Schließlich muß, um die intensive Beschäfti-

1) W. ZIFREUND: "Über den Zusammenhang von Programmierter Instruktion, Unterrichtstechnologie und Unterrichtsforschung", in: pl 1, 1969

2) Vgl. H.E. SILBERMAN: "Trends in der Erziehungstechnologie", in: L.J. ISSING: "Der Programmierte Unterricht in den USA heute", Weinheim 1967

gung mit der Programmierten Unterweisung verständlich zu machen, noch darauf hingewiesen werden, daß sich Lehrprogramme in vorzüglicher Weise für die experimentelle Erforschung von Lehreinheiten verschiedener Struktur eignen, da - im Gegensatz zu allen herkömmlichen Unterrichtsmethoden - die experimentellen Bedingungen, soweit es sich um die unabhängigen Variablen handelt, relativ eindeutig definiert und vor allem kontrolliert werden können. Da der Versuch einer rationaleren Gliederung pädagogischer Maßnahmen der experimentellen Aufhellung von Unterrichtsstrukturen bedarf, ist der Rückgriff auf die Programmierte Unterweisung im Rahmen der Unterrichtsforschung unerlässlich.

Im folgenden wollen wir einen allgemeinen Überblick über wesentliche Arbeiten und methodische Ansätze im Rahmen der Programmierten Unterweisung geben. Angesichts der kaum mehr übersehbaren Literatur auf diesem Gebiet versuchen wir einen "repräsentativen Ausschnitt" darzustellen, jeder weitergehende Versuch würde uns in Richtung einer Bibliographie führen und damit den Rahmen und die Absicht dieser Arbeit sprengen. Der Versuch, einen "repräsentativen Ausschnitt" darzustellen, scheint uns unerlässlich zu sein, da die Zielsetzung dieser Arbeit nur innerhalb des Kontextes der gesamten einschlägigen Forschungsbemühungen sinnvoll eingeordnet werden kann.

2.10 Vergleichsstudien

Hierunter fallen alle Untersuchungen, welche die Programmierte Unterweisung mit herkömmlichen Unterrichtsmethoden vergleichen. Dabei stellen die verglichenen Unterrichtsmethoden die unabhängigen Variablen dar, als abhängige Variablen werden Lernleistung, Arbeitszeit, Einstellung zum Unterricht u.a. herangezogen. Wir werden der Diskussion des Forschungsstandes gerade auf diesem Gebiet besonders viel Raum widmen und über eine allgemeine Übersicht hinaus kritisch kommentieren, da unsere Arbeit genau an dieser Stelle ansetzt.

Die harte Kritik SKINNERS am herkömmlichen Unterricht und die von ihm im Rahmen seiner Lerntheorie behauptete Überlegenheit der Programmierten Unterweisung forderte eine experimentelle Überprüfung geradezu heraus. Dazu kommt, daß seitens der Schulen bzw. der industriellen Ausbildung ein großes Interesse daran besteht, die Wirksamkeit der neuen Lehrmethode vergleichend mit herkömmlichen Unterrichtsverfahren zu überprüfen. Auf diesem Hintergrund entstand schon relativ früh eine Vielzahl von Arbeiten, in welchen versucht wurde, den Lehreffekt der neuen Lehrmethode mit demjenigen des herkömmlichen Unterrichts zu vergleichen. An dieser Stelle seien stellvertretend die Veröffentlichungen von E. KEISLAR¹⁾, J. EVANS²⁾,

1) E. KEISLAR: "The Development of Understanding in Arithmetic by a Teaching Machine", in: J.P. De CECCO (Hrsg.): "Educational Technology", New York 1964

2) J.L. EVANS: "Programing in Mathematics and Logic", in: R. GLASER (Hrsg.): "Teaching Machines and Programed Learning II", New York 1965

R. INGLE and W. GEPHARDT¹⁾ und D. WALLIS and R.P. WICKS²⁾ genannt. Im deutschsprachigen Bereich sind insbesondere die Arbeiten von R. GLASER, J.H. REYNOLDS und M.G. FULLICK³⁾ bekanntgeworden sowie die von W. SCHRAMM⁴⁾ berichteten Untersuchungen. Bei L.J. ISSING⁵⁾ finden sich 50 einschlägige Arbeiten aus dem angelsächsischen Bereich allein für den Zeitraum von 1955 - 1966, wobei die Vielzahl der nichtveröffentlichten, z.T. nur hektographiert in wenigen Exemplaren kursierenden, Arbeiten nicht mit berücksichtigt ist. Relativ wenig Studien liegen dagegen im mitteleuropäischen, insbesondere im deutschsprachigen Raum vor. Bekanntgeworden sind hier vor allem

-
- 1) R.B. INGLE and W.J. GEPHARDT: "A Critique of a Research Report: Programmes Instruction Versus Usual Classroom Procedures in Teaching Boys to Read", in: Am. Educ. Research Journal, Bd. III, I, Jan. 1966
 - 2) D. WALLIS and R.P. WICKS: "The Autotutor and Classroom Instruction: Three Comparative Studies. 1. The Royal Navy Study, in: Occupational Psychology, 1963, 37
 - 3) R. GLASER, J.H. REYNOLDS, M.G. FULLICK: "Programmierter Unterricht in der Klasse", in: W. CORRELL (Hrsg.): "Programmierter Unterricht und Lehrmaschinen", Braunschweig 1966
 - 4) W. SCHRAMM (Hrsg.): "Vier Fallstudien über programmierten Unterricht", Berlin - Bielefeld 1966
 - 5) L.J. ISSING: "Forschung und Forschungsergebnisse im Bereich des Programmierten Unterrichts - eine bibliographische Übersicht", in: L.J. ISSING (Hrsg.): "Der Programmierte Unterricht in den USA heute", Weinheim 1967

die Arbeiten von G. HUBER¹⁾, W. HOCHHEIMER²⁾ und K. RÖTTEL³⁾. Eine nähere Analyse der deutschsprachigen Literatur zeigt allerdings, daß es eine Vielzahl von sog. "Erfahrungsberichten" gibt, in welchen die Programmierte Unterweisung mit dem herkömmlichen Unterricht verglichen wird⁴⁾. Diese Arbeiten greifen meistens auf sehr kleine Stichproben und kurze Programmeinheiten zurück. Selten sind Erfahrungsberichte, welche mit großen Stichproben und langen Programmsequenzen arbeiten⁵⁾. Die überwiegende Mehrzahl dieser Erfahrungsberichte haben allerdings keine wissenschaftliche Aussagekraft, da sie entweder methodisch unzu-

-
- 1) G. HUBER: "Lernpsychologische Befunde bei Programmierter Unterweisung", München 1966
 - 2) W. HOCHHEIMER: "Untersuchungen und Erprobungen zum Programmieren Lernen", in: ZeF 1, 1967, 1
 - 3) K. RÖTTEL: "Vergleichende Analyse einer Versuchsserie mit programmiertem und konventionellem Unterricht", in: ZeF, 2, 1968, 4
 - 4) An dieser Stelle seien nur die Arbeiten von S. MIZERA und B. FISNER: "Ein Versuch mit dem programmierten Unterricht der Chemie an der Maschinenbauakultät der Tschechischen Technischen Hochschule in Prag", in: pl 2, 1968; F.W. SCHÜTT: "Unterrichtsversuche im Chemieunterricht des Gymnasiums bei programmiertem und konventionellem Unterricht", in: pl 3 (1968) und H. LUSCHER: "Programmierter Unterricht im Spiegel erster Erfahrungen", Braunschweig 1964, erwähnt.
 - 5) O. HARDE: "Ein Jahrgangsprogramm für den Rechenunterricht", in: B. ROLLETT (Hrsg.): "Praxis und Theorie des Programmierten Unterrichtes", Stuttgart-München 1970 und W. DRALLE: "Erfahrungen mit einem Jahrgangsprogramm", in: B. ROLLETT: "Praxis und Theorie des Programmierten Unterrichtes", Stuttgart-München 1970

länglich angesetzt sind und z.T. völlig unzureichende Verrechnungstechniken aufweisen. Allen Vergleichsstudien gemeinsam ist die Fragestellung, ob Schüler, die programmiert arbeiten, überhaupt einen Lerngewinn haben und ob dieser Lerngewinn kleiner oder größer ist als bei anderen Unterrichtsverfahren, wobei letztere überwiegend durch das sog. frontalunterrichtliche Verfahren repräsentiert werden. Dieses wird meist ziemlich unexakt als "konventioneller Unterricht" umschrieben. Weiter ist für die meisten Vergleichsstudien kennzeichnend, daß sie feldexperimentell angelegt sind, d.h. es wird auf "geschlossene" Gruppen, meistens Klassen, innerhalb der Schulen zurückgegriffen. Hier liegt ein grundsätzlicher Unterschied zu den üblichen lernpsychologischen Untersuchungen vor, welche fast ausschließlich im Labor angesetzt werden und randomisierte Stichproben aufweisen. Wir werden auf diese Problematik unten noch ausführlich zu sprechen kommen. Der Rückgriff auf natürliche Klasseneinheiten wird, wenn überhaupt, im Hinblick auf die externe Validität gerechtfertigt. Die überwiegende Mehrzahl dieser Vergleichsstudien erbrachte nicht signifikante Differenzen zwischen der Lehreffektivität der verglichenen Unterrichtsmethoden, in einigen Fällen wurde von einer signifikant größeren Lehreffektivität der Programmierten Unterweisung berichtet, in sehr wenigen Fällen erwies sich der herkömmliche Unterricht als überlegen. Bei den oben erwähnten Erfahrungsberichten sind die Ergebnisse etwas eindeutiger. Hier wird sehr häufig von einer Überlegenheit der Programmierten Unterweisung berichtet, Ursache dafür sind zweifellos die fehlenden Signifikanzberechnungen und der Rückgriff auf unzulässig interpretierte Prozentwerte, eine nähere Analyse der angegebenen Zahlenwerte weist im allgemeinen deutlich auf nicht signifi-

kante Differenzen hin. Zusammenfassend muß davon ausgegangen werden, daß es den meisten der bisher durchgeführten Vergleichsstudien nicht gelang, eine signifikant höhere Lehreffektivität der Programmierten Unterweisung überzeugend nachzuweisen. Ein solches Ergebnis ist zunächst überraschend, da aufgrund der theoretischen Analyse der verglichenen Unterrichtsmethoden eine grundsätzliche Überlegenheit der Programmierten Unterweisung erwartet werden müßte und von kompetenter Seite (SKINNER) auch nachdrücklich vertreten wurde.

Daß die meisten Untersucher mit dieser Erwartung an ihre Arbeit herangingen, zeigen die Formulierungen der Alternativhypothesen, welche in der überwiegenden Anzahl eine Überlegenheit der Programmierten Unterweisung beinhalten; zweiseitige Fragestellungen sind äußerst selten. Eine nähere Durchsicht der einschlägigen Arbeiten läßt allerdings vermuten, daß die nicht signifikanten Unterschiede zum Teil Folge deutlich erkennbarer experimenteller Mängel sind. Wir werden in unserem methodischen Teil ausführlich darauf eingehen.

Folge der unbefriedigenden Ergebnisse war, daß die wissenschaftliche Aussagekraft derartiger Vergleichsuntersuchungen häufig bezweifelt wurde, zum Teil von Autoren, welche selbst entsprechende Arbeiten veröffentlicht hatten. So äußern sich u.a. J.L. EVANS, R. GLASER und L.E. HOMME¹⁾ und I.G. GOTKIN²⁾ sehr kritisch. "Kein einzi-

1) Vgl. J.L. EVANS, R. GLASER und L.E. HOMME: "A Preliminary Investigation in the Properties of Verbal Learning Sequences of the 'Teaching Machines' Type", in: R. GLASER und A.A. LUMSDAINE (Ed.): "Programmed Learning and Teaching Machines, New York 1960, S. 446-451

2) I.G. GOTKIN: "Experimentation with Programed Instruction: Some Suggestions for Educators", in: Programed Instruction, 1,1, 1962, S. 29-36

ges Programm kann "die neue Methode" repräsentieren und kein Lehrer "den Durchschnittslehrer". Vergleiche zwischen einem Lehrer und einem Programm haben wahrscheinlich deshalb keinen Sinn, weil wir bis heute weder für den Lehrer noch für den Programmierten Unterricht brauchbare Bewertungsmaßstäbe besitzen"¹⁾. So sehr dieses Argument im Hinblick auf die Praxis der meisten Vergleichsuntersuchungen zutrifft - und von hierher muß es nach unserer Auffassung auch verstanden werden -, in welchen auf der einen Seite kaum oder überhaupt kein Bemühen erkennbar ist, den herkömmlichen Unterricht zu definieren und auf der anderen Seite von d e r Programmierten Unterweisung geredet wird, obwohl den Untersuchungen meist nur ein Programm zugrunde liegt, und aus deren Ergebnisformulierungen ein völlig unkritischer Generalisierungsanspruch entnommen werden muß, so sehr muß es im Kontext einer methodenkritischen Diskussion mit Vorsicht aufgenommen werden. Keine vernünftig angelegte Untersuchung kann zum gegenwärtigen Zeitpunkt für sich den Anspruch erheben, d i e Programmierte Unterweisung mit d e m konventionellen Unterricht vergleichen zu wollen. Uns scheinen hier die Erwartungen, welche Vergleichsstudien entgegengebracht werden, viel zu hoch angesetzt zu sein. Dies äußert sich, wie wir oben bereits erwähnten, in den einschlägigen Arbeiten häufig in unzulässigen Generalisierungen. Wenn Vergleichsuntersuchungen angesetzt werden, kann vernünftigerweise zunächst nur davon ausgegangen werden, daß e i n Programm mit e i n e m ganz bestimmten herkömmlichen Unterricht verglichen wird. Sofern sich Programm und herkömmlicher Unterricht durch

1) L.M. STOLUROW: "A Response to 'Inside Opinion' ", in: Programed Instruction, 1,2, 1962, S. 2

klar definierte Merkmale beschreiben lassen, was beim Programm immer der Fall ist und beim herkömmlichen Unterricht in einer sauber angelegten Untersuchung immer der Fall sein sollte, können die Ergebnisse der Untersuchung hypothetisch in einen größeren Zusammenhang eingeordnet werden, ohne daß zunächst eine Übergeneralisierung vorliegt. Allerdings sollten gerade auf diesem Gebiet der Unterrichtsforschung wenigstens partielle Wiederholungsuntersuchungen angesetzt werden, bevor Untersuchungsergebnisse in den Kontext der theoretischen Diskussion aufgenommen werden. "Bevor im Bereich des Programmierten Unterrichts keine Wiederholung von Experimenten stattfindet, haben wir nur geringe Verlässlichkeit, und demzufolge werden wir bei zukünftigen Untersuchungen behindert sein, da wir durch nicht wiederholbare Experimente irregeführt werden können"¹⁾. An dieser Stelle muß unter Bezugnahme auf STOLUROWS Kritik an Vergleichsuntersuchungen ausdrücklich vermerkt werden, daß diese streng genommen keine Aussage über den Unterschied zwischen den verglichenen Methoden bezüglich Stoffstrukturierung, Motivierungshilfen, Verstärkungstechnik u.a. machen wollen, sondern zunächst lediglich die Frage nach dem Unterschied der Lehreffektivität derselben gestellt wird. Deswegen scheint es uns durchaus angemessen zu sein, die verglichenen Lehrmethoden mit Hilfe möglichst konstanter, valider und operational beschreibbarer Merkmale zu definieren und den nicht definierten (nicht definierbaren) Anteil bezüglich der Auswirkung auf die abhängige Variable "Lehreffektivität" als unsystematische Störvarianz zu betrachten bzw. auf ein experimental design zurückzu-

1) L.D. EIGEN: "Probleme der Forschung im Bereich des Programmierten Unterrichts", in: L.J. ISSING (Hrsg.): "Die Programmierte Unterweisung in den USA heute", Weinheim 1967, S. 84

greifen, welches die Störvarianz möglichst weitgehend unter Kontrolle hält. Sicher gilt dann immer noch das Argument, daß der herkömmliche Unterricht weniger präzise zu beschreiben ist als das Vergleichsprogramm; wir meinen uns allerdings hier dem häufig geäußerten Argument anschließen zu können, daß unbesehen der Abweichung der verschiedenen "konventionellen Unterrichte" unter bestimmten - von uns noch zu diskutierenden - Bedingungen von einer durchschnittlichen Lehreffektivität ausgegangen werden kann¹⁾. Das Festhalten an - im Sinne der klassischen Experimentalpsychologie - völlig eindeutig definierten Methoden hätte zur Konsequenz, daß ein Großteil der Fragestellungen innerhalb der gesamten gegenwärtigen Unterrichtsforschung als "sinnlos" bezeichnet werden müßte, da sich hier genau dieselben oder jedenfalls sehr ähnliche methodische Probleme stellen. So wären z.B. Untersuchungen über die Auswirkung von Erziehungsstilen auf das Schülerverhalten ohne zwingende wissenschaftliche Aussagekraft, da es nicht den autoritären und den demokratischen Unterricht gibt, sondern eben nur eine Vielzahl von autoritären bzw. demokratischen Erziehungsstilen, welche auf der einen Seite jeweils gemeinsame und ziemlich eindeutig definierbare Eigenschaften haben, auf der anderen Seite jedoch auch in vielerlei Hinsicht variieren und nur unzulänglich beschreibbar sind. Kritisch wird es allerdings, wenn gefragt wird, warum eine Unterrichtsmethode einer anderen überlegen ist, sofern hinter solch einer Frage der Versuch steht zu beweisen, daß jeweils zu definierende Struktureigentümlichkeiten bzw. strukturelle Differenzen der verglichenen Unterrichtsmethoden einen höheren bzw. niedrigeren

1) Einen sehr interessanten methodischen Beitrag lieferte in diesem Zusammenhang K. ECKEL: "Die Bedeutung des Klasseneffekts für die Schulpädagogische Forschung", in: pl 3, 1969

Lehrerfolg bewirken. Im Rahmen von Vergleichsstudien kann bei Vorliegen signifikanter Lerndifferenzen zunächst nur hypothetisch gesagt werden, daß die definierten Eigenschaften der verglichenen Unterrichtsmethoden (z.B. im Falle der Programmierten Unterweisung die häufige und regelmäßige Verstärkung bzw. deren Fehlen im herkömmlichen Unterricht) die Differenz bewirken. Tatsächlich liegen nun Hinweise vor, daß die die verglichenen Unterrichtsmethoden kennzeichnenden Eigenschaften, welche den Hypothesen bislang zugrunde gelegt wurden, keine geeigneten Beurteilungsmaßstäbe für die Begründung der Lehreffektivität einer Unterrichtsmethode darstellen (vgl. 2.20). Hier wird eine weitere, über Vergleichsstudien hinausgehende, experimentelle Klärung ansetzen müssen, wobei in diesem Fall strukturelle Eigentümlichkeiten innerhalb einer Methode variiert werden müssen¹⁾. Solche Untersuchungen liegen zwar schon in großer Zahl vor, deren Ergebnisse sind aber widersprüchlich und insgesamt unbefriedigend.

Wir wollen an dieser Stelle noch einmal festhalten, daß in Vergleichsstudien zunächst lediglich die Frage geklärt werden kann, ob ein bestimmtes Programm frontalunterrichtlich arbeitenden, durchschnittlich qualifizierten Lehrern bezüglich der Lehreffektivität überlegen ist und im weiteren, ob ein bestimmter Programmtyp, der einen bestimmten Stoffinhalt vermittelt, dem frontalunterrichtlichen Vorgehen eines beliebigen durchschnittlich qualifizierten Lehrers bezüglich der Lehreffektivität überlegen ist. Die experimentelle Klärung

1) Entsprechende Probleme stellen sich im Rahmen der Erforschung der sog. Erziehungsstile. Auch hier liegen bereits Arbeiten vor, welche strukturelle Eigentümlichkeiten der untersuchten Erziehungsstile variieren. Vgl. K. EYFERT: "Methoden zur Erfassung von Erziehungsstilen", in: Th. HERMANN (Hrsg.): "Psychologie der Erziehungsstile", Göttingen 1966, S. 20 ff.

solcher Fragen ist für die Schulpraxis zweifellos von wesentlichem Belang, will man sie nicht zwingen, auf ihre nachgewiesenermaßen unqualifizierten "qualitativen" Beurteilungskriterien zurückzugreifen¹⁾. Es ist diejenige Frage, welche Lehrer zu allererst an die programmierte Unterweisung stellen, wobei sie i.allg. negativ formuliert wird, d.h. es wird seitens der Schulen zunächst davon ausgegangen, daß eine neue Methode möglicherweise schlechter als die bislang praktizierten ist. Darüber hinaus scheinen uns Vergleichsstudien, sofern sie methodisch sauber angesetzt sind und - mit gleichem Ergebnis - (partiell oder ganz) wiederholt werden, Ansatzpunkte für differenziertere und weiterführende Forschungsarbeiten zu sein, welche nach der oder den Ursachen der Überlegenheit bzw. Unterlegenheit einer Unterrichtsmethode fragen. Angesichts der weithin nicht-signifikanten Ergebnisse der bisherigen Vergleichsstudien könnte man allerdings folgern, daß sie sich zur Klärung der oben ausgeführten Fragen als untauglich erwiesen haben und deswegen fallen gelassen werden sollten. Wir haben aber bereits oben darauf hingewiesen und werden es im methodischen Teil in aller Ausführlichkeit tun, daß die nichtsignifikanten Ergebnisse häufig auf die unzureichende Kontrolle der experimentellen Situation und d.h. letztlich auf unzureichende experimental designs zurückgeführt werden können. Hier klingt bereits an, daß sich die Schwierigkeiten der Vergleichsstudien in weiten Bereichen mit den methodischen Problemen der experimentellen Feldstudien innerhalb der Unterrichtsforschung decken und nicht etwa von einer falsch angesetzten Fragerichtung her erklärt werden können.

1) Vgl. E.Z. ROTHKOPF: "The Epidemiology of Ignorance", in: Contemp. Psychol., 9, 1964, S. 234-236

Angesichts der Schwierigkeiten, den herkömmlichen Unterricht zu definieren, schlug K.J. KLAUER¹⁾ vor, sowohl Versuchs- als auch Kontrollgruppe einer "eindeutigen" Lernsituation auszusetzen. So könnte z.B. die Versuchsgruppe nach einem Vortest einem mathematischen Lehrprogramm ausgesetzt werden und anschließend den Schlußtest absolvieren, währenddessen die Kontrollgruppe nach demselben Vortest ein historisches Lehrprogramm durcharbeitet und sich anschließend dem Schlußtest der ersten Gruppe unterzieht. Mit Hilfe einer solchen Versuchsanordnung kann festgestellt werden, ob das mathematische Lehrprogramm überhaupt mathematisches Lernen provoziert. Eine solche Fragestellung scheint uns jedoch wenig angemessen zu sein, da kein vernünftiger Grund zu der Annahme besteht, daß ein mathematisches Programm keine mathematischen Lernprozesse provoziert. Die weitergehende und notwendige Frage, in welchem Ausmaß mathematisches Wissen bewirkt wird, kann nur unzureichend geklärt werden, da von vornherein angenommen werden muß, daß die Versuchsgruppe im Schlußtest signifikant besser ist als die Kontrollgruppe, welche sich ja während der experimentellen Phase mit einem völlig anderen Lehrstoff befaßt hat. Bei einem solchen Ansatz muß auch ein schlechtes Programm als valide angesehen werden, zudem wird die Frage, ob der Lehrer anstelle des Programms unterrichten soll, nicht geklärt. Hier wird, um methodische und definitorische Schwierigkeiten zu vermeiden, ein methodisch zwar exakterer, aber bezüglich seiner Aussagekraft trivialer Ansatz vorgeschlagen, wobei zusätzlich darauf hingewiesen werden muß, daß der Autor davon ausgeht, daß ein Programm - im Gegensatz zum herkömmli-

1) Vgl. K.J. KLAUER: "Über die Erfassung der Validität von Lehrprogrammen", in: pl, 1, 1968, S. 2

chen Unterricht - bezüglich der strukturellen Eigentümlichkeiten, welche seine Lehreffektivität bewirken, "eindeutig" definierbar ist. Wenn schon der Schwerpunkt der Diskussion auf die Frage der Definierbarkeit der unabhängigen Variablen gelegt wird, sollte man davon ausgehen, daß ein Programm zwar exakt beschreibbar ist, daß damit aber noch längst nicht die Frage geklärt ist, ob die Beschreibungskriterien mit der Lehreffektivität des Programms zusammenhängen und in welchem Ausmaß dies der Fall ist. Im übrigen halten wir es für angemessen, daß in die Diskussion der Möglichkeiten von Vergleichsstudien die Probleme optimaler experimental designs und der entsprechenden Verrechnungstechniken¹⁾ aufgenommen werden sollte.

Anders ist es zu werten, wenn die Kontrollgruppe ein Lehrprogramm mit demselben Inhalt durcharbeitet, wobei sich das Programm der Kontrollgruppe durch jeweils eindeutig definierte strukturelle Eigentümlichkeiten vom Programm der Versuchsgruppe abheben sollte. In dieser Richtung setzten insbesondere in den Vereinigten Staaten viele Autoren Untersuchungen an²⁾; wir werden unter 2.20 darauf zurückkommen. An dieser Stelle sei nur vermerkt, daß solche Arbeiten zwar Auskunft darüber geben können, welches von zwei bzw. mehreren Programmen bessere Lerneffekte bewirkt und vor allem wodurch die höhere Lehreffektivität hervorgerufen wird, die für den Schulpraktiker entscheidende Frage, ob der Einsatz programmierter Lernmaterials unter Verzicht auf den herkömmlichen Unterricht vertretbar ist, bleibt aber völlig ungeklärt.

1) D.T. CAMPBELL and J.C. STANLEY: "Experimental and Quasi-Experimental Designs for Research on Teaching", in: N.L. GAGE (Ed.): "Handbook of Research on Teaching", Chicago 1965

2) Vgl. L.J. ISSING, a.a.O., S. 102 ff.

Eine neue Lehrmethode kann aber nach unserer Auffassung in der Schulpraxis erst dann eingesetzt werden, wenn der Nachweis erbracht ist, daß sie bezüglich der erwünschten Lehreffektivität mindestens genauso gut ist wie dasjenige Unterrichtsverfahren, welches sie ersetzen soll. Diese Frage muß im Hinblick auf alle schulisch relevanten Stoffgebiete und alle grundlegenden Programmformen geklärt werden. Eine Analyse der einschlägigen Forschungsbemühungen in jüngster Zeit zeigt deutlich, daß die Frage nach der Lehreffektivität im Zentrum der Betrachtung steht und das Problem der Verursachung derselben in den Hintergrund drängt. In diesem Sinne ist auch G. EIGLER zu verstehen, wenn er schreibt: "Der gegenwärtige programmierte Unterricht ist pragmatisch, insofern nur der Lehrerfolg als Qualifikationskriterium gilt"¹⁾. Daß die gegenwärtige Forschung auf die Definition von "brauchbaren Bewertungsmaßstäben", abgesehen von dem Kriterium des Lehrerfolgs, weitgehend verzichtet, geht aus der oben erwähnten pragmatischen Orientierung der gegenwärtigen einschlägigen Forschung hervor.

" ... Vom Projekt einer wissenschaftlichen Organisation der Lehr- und Lernprozesse ist wissenschaftlich derzeit nur noch die Kontrolle des Lehrerfolgs übrig geblieben^{2,3)}.

1) G. EIGLER: "Prüfung von Lehrprogrammen", in: ZeF, 2, 1968, 3, S. 122

2) G. EIGLER; a.a.O., S. 122

3) Der Versuch von M. GAGNÉ (vgl. dazu M. GAGNÉ: "The analysis of instructional objectives for the design of instruction", in: R. GLASER (Ed.): "Teaching machines and programmed learning, II", New York 1965, S. 21 ff.), unterschiedliche Modelle des Lehrens im Hinblick auf Lehrziele verschiedener Struktur zu entwerfen, deuten zwar eine andere Richtung an, sie werden aber die pragmatische Orientierung bestimmter Bereiche der gegenwärtigen Unterrichtsforschung nicht verändern sondern nur ergänzen können.

2.20 Experimentelle Analyse der Variablen, welche den Lehreffekt der Programmierten Unterweisung beeinflussen

Im folgenden soll aufgezeigt werden, daß auch bei Verzicht auf die schwer definierbare unabhängige Variable "herkömmlicher Unterricht" über die Vergleichsstudien hinaus sich bei fast allen Forschungsansätzen im Bereich der Programmierten Unterweisung die Probleme nicht signifikanter und widersprüchlicher Ergebnisse und einer z.T. extrem eingeschränkten Generalisierbarkeit der Befunde stellen. Besondere Aufmerksamkeit werden wir dabei dem Problem der Verstärkung von Schüleraktivitäten und dem Prinzip der aktiven Antwortformulierung zuwenden.

Wir haben unter 2.10 darauf hingewiesen, daß KLAUER den Vergleich zweier bzw. mehrerer Programme, welche sich bezüglich definierter Konstruktionsprinzipien unterscheiden, für wissenschaftlich besonders ergiebig hält. Begründet wird dies einmal damit, daß die verglichenen Lehrmethoden schärfer definiert seien als im Falle eines Vergleichs mit einem herkömmlichen Unterrichtsverfahren, zum anderen wird der sog. Hawthorne-Effekt ausgeschaltet, welcher bei herkömmlichen Vergleichsstudien i.allg. die programmiert unterrichtete Gruppe begünstigt und als systematische Störvarianz in die Ergebnisse mit eingeht. Für die Entwicklung der optimalen didaktischen Stoffaufbereitung von Unterrichtseinheiten scheint uns dieser Ansatz vielversprechend zu sein. Hier bietet sich zum ersten Mal die Möglichkeit, voneinander abweichende didaktische Mikrostrukturen einem sauberen experimentellen - und das heißt in diesem Zusammenhang insbesondere wiederholbaren - Vergleich zu unterziehen. Tatsächlich wendete sich ein großer Teil der Forschung in den sechziger Jahren solchen Studien zu, wobei auch hier

insbesondere in den Vereinigten Staaten Veröffentlichungen vorliegen. Im deutschsprachigen Raum blieb es größtenteils bei der Konstatierung der Notwendigkeit solcher Untersuchungen. Dies wird u.a. aus der Tatsache ersichtlich, daß in den deutschen Fachzeitschriften¹⁾ während den letzten drei Jahren kaum einschlägige Arbeiten veröffentlicht worden sind. Dies mag z.T. daran liegen, daß die Beschäftigung mit der Programmierten Unterweisung aus dem Bereich der Lernpsychologie herausgenommen worden ist und auf der einen Seite von Schulpraktikern bearbeitet wurde, welche meist unzulängliche, in keiner Weise repräsentative Erfahrungsberichte lieferten, auf der anderen Seite von kybernetisch orientierten Pädagogen und Sprachwissenschaftlern, welche sich vorwiegend mit Problemen der computerunterstützten Unterweisung und des Sprachlabors befassen. Experimentelle Arbeiten bezüglich der "Erforschung des Lernens unter den verschiedenen Bedingungen der Lehre"²⁾ finden sich daher vorwiegend dort, wo die Erforschung der Programmierten Unterweisung relativ eng mit der Lernpsychologie verbunden ist, d.h. in den Vereinigten Staaten. Diese Verbindung zur Lernpsychologie scheint uns besonders deswegen sehr fruchtbar zu sein, weil sie einen Rückgriff auf eine lange und differenzierte experimentelle Erfahrung ermöglicht.

-
- 1) Dies sind in erster Linie die "Zeitschrift für erziehungswissenschaftliche Forschung", (ZeF), München, und die Zeitschrift "Programmiertes Lernen, Unterrichtstechnologie und Unterrichtsforschung", (pl), Berlin
 - 2) K.J. KLAUER, a.a.O., S. 2

Wir werden im folgenden die einzelnen Forschungsansätze in Anlehnung an L.J. ISSINGS Bibliographie amerikanischer Veröffentlichungen¹⁾ klassifizieren und in der entsprechenden Abfolge berichten.

2.21 Untersuchungen, welche den Darbietungsmodus betreffen:

Hierunter fassen wir diejenigen Arbeiten zusammen, welche die Reihenfolge der Darbietung von Lehreinheiten variieren und im Hinblick auf die Lehreffektivität analysieren.

2.211 Eine Vorgehensweise besteht darin, daß eine Zufallsabfolge von Lernschritten mit einer logischen Abfolge verglichen wurde²⁾. In relativ vielen Untersuchungen ergaben sich dabei nichtsignifikante Unterschiede, die zufällig angeordneten Lehreinheiten riefen nur unbedeutend schlechtere Lernergebnisse hervor als logisch strukturierte. Solche Ergebnisse sind angesichts einer Vielzahl lernpsychologischer und didaktischer Untersuchungen³⁾, deren Ergebnisse in eine völlig andere Richtung weisen, ausgesprochen überraschend. Sie müssen wohl

1) L.J. ISSING, a.a.O., S. 102 ff.

2) In diesem Zusammenhang sei auf folgende Arbeiten verwiesen: E.L. GAVURIN und V.A. DONAHUE: "Logical Sequence and Random Sequence Teaching Machine Programs", in: Automated Teaching Bulletin, I, 1961, S. 3-9; R.F. MAGER: "On the Sequencing of Instructional Content", in: J.P. De GECCO (Ed.): "Educational Technology", New York, 1964, S. 132-142 und A. ROE: "Automated Teaching methods using linear programs", in: Journ. appl. Psychol., 1962, 46, S. 198-201.

3) Vgl. H. SKOWRONEK: "Psychologische Grundlagen einer Didaktik der Denkerziehung", Hannover 1970

vor allem im Hinblick auf die den Untersuchungen zugrunde gelegten Lerninhalten und der ausgesprochen geringen Anzahl an Lernschritten erklärt werden. Wir meinen uns in voller Übereinstimmung mit der gegenwärtigen Didaktik zu befinden, wenn wir konstatieren, daß die Frage nicht lauten kann, ob Lernschritte zufällig oder logisch angeordnet werden sollen, sondern sie fragt vielmehr, welche (in jedem Fall nicht zufällige) Anordnung für einen bestimmten Unterrichtsstoff und eine definierte Adressatengruppe optimal ist¹⁾. Daß sich dabei nicht ausschließlich das Problem der logischen, sondern u.a. auch der altersgemäßen Anordnung und Gestaltung von Lernschritten stellt, ist spätestens seit der didaktischen Auswertung des Systems von J. PIAGET durch AEBLI²⁾ selbstverständlich. An dieser Stelle muß noch vermerkt werden, daß gerade das Konzept der Programmierten Unterweisung eine altersgemäße Stoffaufbereitung und Darbietung anzielt. Dies wird aus den einschlägigen Validierungskonzepten für Unterrichtsprogramme deutlich³⁾. Wenn W. SCHRAMM⁴⁾ z.B. berichtet, daß Schülern die Aufgabe gestellt wurde, "sich ein bestimmtes Wissen dadurch anzueignen, daß sie einen Lehrer von sich aus befragten" und dabei keinerlei Ähnlichkeit mit

-
- 1) Entsprechende Vorschläge finden sich in R. GLASER: "Some research problems in automated instruction: instructional programming and subject-matter structure", in: J.E. COULSON (Ed.): "Programmed learning and computer-based instruction", New York 1962
 - 2) H. AEBLI: "Psychologische Didaktik", Stuttgart 1968
 - 3) Vgl. J.P. LYSAUGHT und C.M. WILLIAMS: "Einführung in die Unterrichtsprogrammierung", München-Wien 1967, S. 51 ff.
 - 4) W. SCHRAMM: "Programmierter Unterricht heute und morgen." Berlin 1963, S. 49

der Arbeitsweise an einem logisch geordneten Programmtext erkennbar war, spricht dies keinesfalls dafür, daß zufällig angeordnete Lernschritte genauso lehrwirksam wie logisch geordnete sind, hier zeigt sich nur, daß die logische Anordnung von Lernschritten nicht der einzige Weg zur optimalen Stoffaufbereitung ist. Auf dieses Problem wies schon WERTHEIMER in einem anderen Zusammenhang hin¹⁾: "Wie Lehrer vorgehen ... ist weitgehend festgelegt durch zwei überlieferte Ansichten über die Natur des Denkens: die Ansicht der traditionellen Logik und die Ansicht der Assoziationstheorie ... aber es ist zumindest eine offene Frage, ob die Art und Weise, in der sie das Denken auffassen, nicht für die eigentliche Denkfähigkeit Hindernisse schafft ...". In jüngster Zeit scheint der Ansatz von E.B. FRY²⁾ erfolgversprechend zu sein, welcher für Mathematik und naturwissenschaftliche Fächer eine logische Gegenstandsordnung, für klassifizierende Wissenschaften eine deskriptive und für die Sozialwissenschaften eine problemzentrierte Anordnung der Lernschritte vorschlägt. Die experimentelle Verifizierung steht allerdings noch aus.

2.212 Eine weitere Untersuchungsrichtung, welche sich mit dem Problem des Darbietungsmodus befaßt, stellt sich die Frage nach der Auswirkung der Größe der Lernschritte auf die Lehreffektivität von Programmen. Diese

1) M. WERTHEIMER: "Produktives Denken". Frankfurt 1957, S. 3

2) Vgl. E.B. FRY: "Teaching Machines and programed instruction. An Introduction", New York 1963, S. 116 ff.

Arbeiten knüpfen alle letztthin an die Forderung von SKINNER nach kleinen Lernschritten an. ISSING weist in seinem zusammenfassenden Bericht¹⁾ zu Recht darauf hin, daß alle einschlägigen Arbeiten darunter leiden, daß es bislang noch nicht gelungen ist, die Größe eines Lernschritts exakt zu definieren. An dieser Stelle seien nur diejenigen Versuche genannt, welche die Größe der Lernschritte mittels der dargebotenen Information²⁾, der Anzahl der verwendeten Worte und dem Schwierigkeitsgrad der verlangten Antwort zu erfassen versuchten. Entsprechend widersprüchlich sind auch die Ergebnisse dieser Arbeiten³⁾. Auch hier ergaben sich häufig nichtsignifikante Unterschiede. Insgesamt kann festgehalten werden, daß sich - wenn überhaupt - Programme mit kurzen Lernschritten als überlegen erwiesen haben, wobei dies indirekt als Bestätigung der SKINNERschen Programmkonzeption aufgefaßt werden kann. In diesem Zusammenhang muß allerdings zugleich einschränkend darauf hingewiesen werden, daß die Auswirkung der Größe der Lernschritte auf die Lernleistung nicht unabhängig von dem vermittelten Stoff und der zugrunde gelegten Schülerstichprobe untersucht werden kann und sich viele der widersprüchlichen Ergebnisse auch von hierher erklären las-

1) Vgl. L.J. ISSING, a.a.O., S. 103

2) Vgl. H. BLANKERTZ: "Theorien und Modelle der Didaktik", München 1969, S. 71 f.

3) In diesem Zusammenhang sei auf die Arbeiten von J.E. COULSON und H.F. SILBERMAN: "Effects of Three Variables in a Teaching Machine", in: Journ. of Educ. Psychol., 51, 1969, S. 135-143 und C.B. SHAY: "Relationship of Intelligence to Step Size on a Teaching Machine Program", in: Journ. of Educ. Psychol., 52, 1961, S. 98-103 verwiesen.

sen. Im deutschsprachigen Raum weist darauf insbesondere SCHIEFELE¹⁾ hin. Interessante Perspektiven eröffnen die Aussagen von COULSON und SILBERMAN²⁾, welche berichten, daß solche Programme eine besonders hohe Lehreffektivität aufweisen, welche den Schülern in gewissen Grenzen ermöglichen, die Aufgabenfolge zu variieren, d.h. z.B. dort kleinere Schritte zu wählen, wo Schwierigkeiten erlebt werden. Interessant ist auch die zusammenfassende Bewertung von ISSING, welcher feststellt, "daß der Lernerfolg dann am größten ist, wenn die Lernschritte während des Lernvorgangs allmählich in ihrer Größe anwachsen"³⁾. Unbesehen dieser Vorstellungen spricht vieles dafür, daß der Versuch, e i n e optimale Lernschrittgröße zu finden, von Anfang an zum Scheitern verurteilt ist, da diese Frage wiederum nur im Hinblick auf das Stoffgebiet und die Adressatengruppe geklärt werden kann.

2.213 Im Zusammenhang mit den Untersuchungen bezüglich der Größe der Lernschritte sind auch jene Arbeiten zu senen, welche den linearen Programmtypus nach SKINNER mit CROWDERS verzweigtem Programmtypus vergleichen⁴⁾.

1) Vgl. H. SCHIEFELE: "Programmierte Unterweisung", München 1964, S. 107

2) J.E. COULSON und H.F. SILBERMAN: "Results of initial experiment in automated teaching", in: A.A. LUMSDAINE and R. GLASER (Ed.): "Teaching machines and programmed learning", Washington 1960, S. 452-468

3) Vgl. L.J. ISSING, a.a.O., S. 103

4) In diesem Zusammenhang sei auf die Arbeiten von A. ROE: "A Comparison of Branching Methods for Programed Learning", in: Journ. of Educ. Res., 55, 1962, S. 407-416 und P.R. WENDT und G. RUST: "Pictorial and Performance Frames in Branching Programed Instruction", in: Journ. of Educ. Res., 55, 1962, S. 430-432 verwiesen.

Auch bei dieser Untersuchungsrichtung zeigt sich, daß von einer generellen Überlegenheit eines Programmtypus nicht gesprochen werden kann. Signifikante Unterschiede der Lehreffektivität sind nicht die Regel, sofern sie vorliegen, finden sie sich in beiden Richtungen. Wie bei fast allen Untersuchungen auf dem Gebiet der Programmierten Unterweisung gilt auch hier der bereits oben mehrfach geäußerte Grundsatz, daß sich solche Fragen nur im Hinblick auf das unterrichtete Stoffgebiet und die angesprochene Adressatengruppe entscheiden lassen.

2.214 Schließlich sei im Zusammenhang mit den Untersuchungen über den Darbietungsmodus noch auf die Vielzahl von Arbeiten hingewiesen, die sich mit der Lehreffektivität von Programmen befassen, welche nach verschiedenen Programmierungstechniken entwickelt worden sind. Da die jeweiligen Techniken z.T. eigens im Hinblick auf spezielle Stoffgebiete entwickelt worden sind und meistens keine gemeinsame theoretische Basis zu anderen Programmierungstechniken vorliegt, können die vorliegenden Bewährungskontrollen nicht oder nur unzulänglich generalisiert werden. ISSING¹⁾ erwähnt allein für den Zeitraum von 1961 bis 1965 24 solche Arbeiten.

1) Vgl. L.J. ISSING, a.a.O., S. 104

2.22 Untersuchungen, welche das Darbietungsmittel betreffen:

Hierunter fallen alle Untersuchungen, welche die Lehr-effektivität zweier oder mehrerer Präsentationsmodi bzw. "medienorganisatorischer Varianten derselben"¹⁾ untersuchen, wobei jeweils dasselbe Lehrprogramm zugrunde gelegt wird. Häufig wurde in solchen Untersuchungen neben Präsentationsmodi wie Buchprogramm und Lehrmaschine als unabhängige Variable der konventionellen Lehrerunterricht mit aufgenommen²⁾. Die meisten Fragestellungen dieser Arbeiten beziehen sich auf praktische Probleme der Schulorganisation und legen ihren Versuchsplänen in Gegensatz zu den unter 2.21, 2.23 und 2.24 berichteten i.allg. natürlichen Klasseneinheiten zugrunde. Besonders beliebt ist der Vergleich der Lehr-effektivität programmierter Textbücher mit maschinellen Darbietungsmitteln, darüber hinaus werden als unabhängige Variablen auch Kriterien wie individuelle Schülerkabine versus Darbietung in üblichen Klassenräumen bzw. Darbietung in Form von Hausaufgaben gewählt. Seltener wurde die Lehreffektivität von akustischer Darbietung versus visueller Präsentation untersucht. Schließlich muß noch die Untersuchungsrichtung individuell gesteuerte Präsentation versus extern gesteuerter Präsentation angeführt werden. Detaillierte Berichte über die einschlägigen Arbeiten finden sich in der Veröffentlichung von STRITTMATTER³⁾. Eine nähere Analyse dieser Arbeiten zeigt, daß "ein Vergleich dieser Untersuchungen hinsichtlich Methode und

1) Vgl. P. STRITTMATTER: "Präsentationsmodi von Lehrprogrammen", Weinheim 1970, S. 4

2) Vgl. D. WALLIS und R.P. WICKS, a.a.O.

3) P. STRITTMATTER, a.a.O., S. 8-92

Ergebnisse nicht möglich"¹⁾ ist, da dafür die methodologischen Voraussetzungen fehlen. Weiter ergeben sich hinsichtlich der Lehreffektivität - entsprechend den unter 2.10 und 2.21 geschilderten Ergebnistendenzen - sehr häufig nichtsignifikante Ergebnisse, die relativ seltenen signifikanten Unterschiede sind ebenfalls widersprüchlich²⁾. Daraus muß die Schlußfolgerung gezogen werden, daß die Generalisierungsmöglichkeit auch bei solchen Arbeiten ebenfalls sehr gering ist, obwohl die unabhängigen Variablen durch die Beschränkung des Vergleichs auf programmierte Unterrichtsformen insofern, als sie Wiederholungsstudien zulassen, strenger definiert sind als der herkömmliche Unterricht in den unter 2.10 berichteten Untersuchungen. "Man wird also immer nur in bezug auf das jeweilige Lehrprogramm und die jeweils überprüfte Adressatengruppe argumentieren bzw. interpretieren können"³⁾. Die hier sichtbar werdende Beschränkung der Aussagekraft entsprechender Arbeiten wird im deutschsprachigen Raum gegenwärtig noch viel zu wenig beachtet.

1) P. STRITTMATER, a.a.O., S. 70

2) Vgl. dazu auch L.J. ISSING, a.a.O., S. 105

3) P. STRITTMATER, a.a.O., S. 92

2.23 Untersuchungen, welche den Antwortmodus betreffen:

Hierunter fallen alle Arbeiten, welche die Auswirkung der Antwortart auf die Lehreffektivität untersuchen. Diese Untersuchungsrichtung geht letztthin auf die Forderung nach ausdrücklichen, selbstproduzierten Antworten zurück, wie sie von SKINNER vertreten wurde. "Der Schüler muß seine Antwort selbst formulieren (im Sinne der Konstruktions-Antwort-Methode), statt sie aus einer Reihe von Auswahlmöglichkeiten zu wählen wie in einem Auswahl-Antwort-Apparat. Ein Grund dafür ist, daß er aus dem Gedächtnis etwas reproduzieren und nicht nur etwas wiedererkennen soll ... Ein anderer Grund dafür ist, daß gutes Auswahl-Antwort-Material auch plausible falsche Antworten enthalten muß, die in dem schwierigen Prozeß der Formung des Verhaltens fehl am Platze sind, weil sie unerwünschte Verhaltensformen verstärken"¹⁾. Entsprechend lautet eine experimentelle Frage, ob Schüler, welche die Antwort selbst produzieren und formulieren, einen höheren Lernerfolg aufweisen als wenn sie aus einer vorgegebenen Anzahl von Lösungsmöglichkeiten eine Antwort auswählen. Hinter dieser Alternativannahme steht i.allg. die Konzeption des verzweigten Programms, wie sie von CROWDER zuerst vorgetragen worden ist. Im Gegensatz zum linearen Programm vom "Typ SKINNER", in welchem "man die Reaktion des Schülers für einen wesentlichen Teil des Lernprozesses"²⁾ hält und "die Frage, wie der Schüler dazu veranlaßt wird, die richtige Ant-

1) Vgl. B.F. SKINNER: "Lehrmaschinen", in: W. CORRELL (Hrsg.): "Programmiertes Lernen und Lehrmaschinen", Braunschweig 1966, S. 41

2) N.A. CROWDER: "Unterschiede zwischen linearem und verzweigtem Programmieren", in: W. CORRELL (Hrsg.): "Zur Theorie und Praxis des programmierten Lernens", Darmstadt 1969, S. 152

wort zu geben, irrelevant ..."¹⁾ ist, geht CROWDER davon aus, daß die Fragen und die daraus resultierenden Antworten primär einen diagnostischen Zweck haben, "und die Grundlage der Technik ist die Tatsache, daß die so gestellte Diagnose sofort verwendet werden kann, um den Schüler mit Hilfsmaterial zu versehen"²⁾. CROWDER meint weiter, daß hinter der Forderung nach aktiver Antwortformulierung und der Überbetonung der Bedeutung der Reaktion für den Lernprozeß die naive und übergeneralisierte Vorstellung des Konditionierungsmodells und dessen unkritische Übertragung auf das menschliche Lernen steht. Für ein linear angelegtes Programm ist es dagegen von entscheidender Bedeutung, daß der Lernende möglichst wenig falsche Reaktionen zeigt. "Bei jeder Diskussion über die Reaktionsweise wird sicherlich die Frage aufgeworfen, ob nicht das Verfahren der Auswahl-Fragen dem Schüler erlaubt, Fehler einzuüben. Wieder ist dieser Punkt für das Abrichten von Tauben relevant, aber mir scheint er irrelevant für die ernsthafte Unterrichtsarbeit mit Menschen zu sein."³⁾ Wir wollen an dieser Stelle außer acht lassen, daß die Behandlung des Problems der Übertragbarkeit des SKINNERschen Konditionierungsmodells auf menschliches Lernen in diesem Zusammenhang von CROWDER nicht gerade sehr tiefgründig diskutiert wird und uns der Frage zuwenden, welche Ergebnistendenzen in den einschlägigen Untersuchungen⁴⁾

1) N.A. CROWDER, a.a.O., S. 152

2) N.A. CROWDER, a.a.O., S. 154

3) N.A. CROWDER, a.a.O., S. 159

4) An dieser Stelle seien die Arbeiten von J.E. COULSON und H.F. SILBERMAN: "Effects of Three Variables in a Teaching Machine", in: Journ. of Educ. Psychol., 51, 1960, S. 135-143 und J.P. WILLIAMS: "Effectiveness and Constructed-Response and Multiple-Choice Programming Modes as a Function of Test Mode", in: Journ. of Educ. Psychol., 56, 1965, S. 111-117 erwähnt.

zu finden sind. Hier wird von FRY¹⁾ berichtet, daß selbst formulierte Antworten dann vorteilhafter sind, wenn das Lernkriterium in einer aktiven Wiedererinnerung besteht, Auswahlantworten dagegen nehmen weniger Zeit in Anspruch. BLYTH²⁾ hält das Auswahlverfahren dann für angebracht, wenn sinnfreie Assoziationen gelernt werden müssen. COULSON und SILBERMAN³⁾ schließlich konnten feststellen, daß mit dem Auswahlverfahren zwar schneller gelernt wird, die Lernunterschiede aber zufällig sind. Insgesamt zeigen sich auch auf diesem Forschungsgebiet sehr häufig nicht signifikante Ergebnisse, soweit sie signifikant sind, sind sie häufig widersprüchlich. Auch hier muß man wohl davon ausgehen, daß entsprechende Fragen nur im Hinblick auf sehr spezielle Stoffgebiete und Adressatengruppen geklärt werden können.

Eine andere experimentelle Fragestellung lautet, ob der Schüler die Antwort ausdrücklich (overt response) formulieren soll, oder ob es genügt, wenn er seine Antwort lediglich denkt (covert response). Wir haben bereits oben geschildert, daß SKINNER und seine Anhänger eine ausdrückliche Antwort für optimal halten, CROWDER und seinen Anhängern genügt die gedachte Antwort. Bei den hier besprochenen Versuchen wurden allerdings lineare Programme zugrunde gelegt. Die Ergebnisse der einschlägigen Untersuchungen sind ebenfalls widersprüchlich.

-
- 1) E.B. FRY: "A study of teaching machine response modes", in: A.A. LUMSDAINE and R. GLASER (Ed.): "Teaching Machines and Programmed Learning", Washington 1960, S. 469-474
 - 2) J.W. BLYTH: "Teaching Machines and human beings", in: A.A. LUMSDAINE and R. GLASER (Ed.): "Teaching Machines and Programmed Learning", Washington 1960, S. 401-415
 - 3) J.E. COULSON and H.F. SILBERMAN: "Results of initial experiment in automated teaching", in: A.A. LUMSDAINE and R. GLASER (Ed.): "Teaching Machines and Programmed Learning", Washington 1960, S. 452-468

So berichten LAMBERT, MILLER und WILEY¹⁾, daß nur gedachte Antworten zeitersparend wirken ohne die Lernleistung herabzusetzen, entsprechende Ergebnisse lieferten ROE²⁾ und STOLUROW und WALKER³⁾. KRUMBOLTZ und WEISMAN⁴⁾ erbrachten dagegen den Nachweis, daß bei ausdrücklichen Antworten die Behaltensleistung signifikant besser ist, ähnliche Ergebnisse finden sich auch bei SCHRAMM⁵⁾. Zusammenfassend kann festgehalten werden, daß in der überwiegenden Anzahl der einschlägigen Untersuchungen keine signifikanten Lernleistungsdifferenzen gefunden werden konnten. Angesichts der Tatsache, daß in den wenigen Veröffentlichungen, in denen signifikante Differenzen berichtet werden, diese zugunsten der ausdrücklichen Antwort ausfallen, kann von einer Tendenz in Richtung auf eine höhere Lerneffektivität bei ausdrücklichen Antworten gesprochen werden. HILGARD⁶⁾ hält diese jedoch für so unbedeutend, daß er feststellt, "daß wenig dafür spricht,

-
- 1) P. LAMBERT, D.M. MILLER and D.E. WILEY: "Experimental folklore and experimentation: The study of programmed learning in the Wauwatosa public Schools", in: Journ. educ. Res., 1962, 55 (9)
 - 2) A. ROE: "Automated teaching methods using linear programs", in: Journ. appl. Psychol., 1962, 46, S. 198-201
 - 3) L.M. STOLUROW and C.C. WALKER: "A comparison of overt and covert response in programmed learning", in: Journ. educ. Res., 1962, 55
 - 4) J.D. KRUMBOLTZ and R.G. WEISMAN: "The effect of overt versus covert responding to programmed instruction on immediate and delayed retention", in: Journ. educ. Psychol., 1962, 53
 - 5) W. SCHRAMM: "Programmierter Unterricht heute und morgen", Berlin-Bielefeld 1963
 - 6) E.R. HILGARD: "Kernprobleme der Lernpsychologie und des programmierten Lernens", in: Die Deutsche Schule, 1963, 10, S. 504

die Antwort hinschreiben zu lassen". In unserem Zusammenhang ist entscheidend, daß die einschlägigen Ergebnisse keinesfalls eindeutig sind und somit diesbezüglich weithin mit den Ergebnissen der Untersuchungen, welche den herkömmlichen Unterricht mit der programmierten Unterweisung vergleichen, "übereinstimmen".

2.24 Untersuchungen, welche die Wirkungsweise der Verstärkung betreffen:

Hierunter fallen alle Arbeiten, welche sich mit der Notwendigkeit und Wirkung der unmittelbaren Verstärkung befassen. Die unmittelbare Verstärkung einer Reaktion ist neben der Technik der Antwortsicherung¹⁾ das wichtigste Verfahren zur Herstellung und Verfestigung des gewünschten Vernaltens, die Verstärkungstheorie bzw. die Theorie des operanten Konditionierens von SKINNER muß sogar als Grundlage der Entwicklung der programmierten Unterweisung überhaupt gesehen werden. "Programmierter Unterricht ist im Labor zum ersten Mal in Form von programmierten Verstärkungszusammenhängen aufgetreten"²⁾. Für SKINNER ist das Fehlen von optimalen Verstärkungsprogrammen, d.h. im Zusammenhang mit der programmierten

1) Eine vergleichende Analyse der Auswirkung dieser beiden Verfahren auf die Lehreffektivität findet sich bei L.M. STOLUROW und H. LIPPERT: "Prompting, Confirmation and Overlearning in the Automated Teaching of a Sight Vocabulary", in: J.P. De CECCO (Ed.): "Educational Technology", New York 1964, S. 187-197. Dabei versteht man unter "Einhilfen" eine Programmierungstechnik, welche Reiz und richtige Antwort gleichzeitig bieten, der Schüler muß die Reiz-Antwortkombination wiederholen. Diese Technik geht auf die Kontinguitätstheorie von GUTHRIE zurück.

2) B.F. SKINNER: "Die Technologie des Lehrens", in: W. CORRELL (Hrsg.): "Zur Theorie und Praxis des programmierten Lernens", Darmstadt 1969, S. 21

Unterweisung das Fehlen von regelmäßiger und sofortiger Verstärkung, die Ursache für das Versagen bzw. die Unzulänglichkeit der herkömmlichen Lehrmethoden. "Die wohl schwerste Kritik gegen den herkömmlichen Schulunterricht betrifft die relative Seltenheit von Verstärkungen. Da der Schüler in seinem Wissen um das Richtige gewöhnlich vom Lehrer abhängig ist und da gewöhnlich viele Schüler von demselben Lehrer abhängen, beträgt die Gesamtzahl der Verstärkungsverbindungen, die etwa in den ersten vier Jahren geschaffen werden kann, kaum mehr als wenige Tausend. Aber nach einer sehr groben Schätzung verlangt ein leistungsfähiges mathematisches Verhalten auf dieser Stufe Verstärkungsmechanismen in einer Größenordnung von 25 000. Wir können vermuten, daß selbst im klügeren Schüler ein bestimmter Verstärkungsmechanismus mehrere Male hergestellt werden muß, ehe er das Verhalten richtig in den Griff bekommen kann."¹⁾ In seinem Artikel "Why we need teaching machines" schreibt SKINNER: "Um zur Kernfrage des Pädagogen zu gelangen, geben wir zunächst einen Überblick über die verfügbaren Verstärker ..." ²⁾ und: "... was an Verstärkungen nötig wäre, um in der gesamten Bevölkerung ein diskriminatives Verhalten zu entwickeln, geht weit über das Leistungsvermögen der Lehrerschaft hinaus ..." ³⁾. Es gibt heute keinen führenden Theoretiker

-
- 1) B.F. SKINNER: "Die Wissenschaft vom Lernen und die Kunst des Lehrens", in: W. CORRELL (Hrsg.): "Programmiertes Lernen und Lehrmaschinen", Braunschweig 1966, S. 74-75
 - 2) B.F. SKINNER: "Warum wir Lehrmaschinen brauchen", in: W. CORRELL (Hrsg.): "Programmiertes Lernen und Lehrmaschinen", Braunschweig 1966, S. 197
 - 3) B.F. SKINNER, a.a.O., S. 200

auf dem Gebiet der Programmierten Unterweisung, der sich mit diesem Problem nicht intensiv befaßt hätte. So schreibt GLASER: "Es scheint gegenwärtig kaum Zweifel darüber zu bestehen, daß die richtige Anwendung von Verstärkungsvorgängen ein bedeutender Aspekt der pädagogischen Technologie sein wird. Das sollte auch so bleiben ... da die zentrale Rolle der Verstärkung bei der Aneignung von Verhalten auf einer lange bekannten empirischen Tatsache beruht"¹⁾. HOLLAND beschreibt die Programmierte Unterweisung wie folgt: "Programmierter Unterricht beruht auf dem einfachen, weithin anerkannten Prinzip, daß ein Verhalten dann gelernt wird, wenn diesem Verhalten bei seinem Auftreten eine Verstärkung folgt. Programmieren bedeutet Anordnung von gezielten Verstärkungen"²⁾. An dieser Stelle muß nun darauf hingewiesen werden, daß ungeachtet der Vielzahl von Vertretern der Bedeutung der regelmäßigen und sofortigen Verstärkung auch Kritiker auf den Plan traten. So schreibt LUMSDAINE: "Ich stelle ernsthaft in Frage, ob Theorien, die den Akzent auf die geschickte Handhabung der Verstärkung legen, ein angemessenes oder auch nur brauchbares theoretisches Modell für die Entwicklung von Lehrprogrammen ... abgeben"³⁾. Allerdings fährt der Autor fort: "Dadurch möchte ich auf keinen Fall den großen Beitrag herabsetzen, der durch die

-
- 1) R. GLASER: "Lernen und Technologie des Unterrichts", in: W. CORRELL (Hrsg.): "Zur Theorie und Praxis des programmierten Unterrichts", Darmstadt 1969, S. 50
 - 2) J.G. HOLLAND: "Reaktionszusammenhänge in Lehrmaschinenprogrammen", in: W. CORRELL (Hrsg.): "Zur Theorie und Praxis des programmierten Lernens", Darmstadt 1969, S. 133
 - 3) A.A. LUMSDAINE: "Probleme beim programmierten Unterricht", in: W. CORRELL (Hrsg.): "Zur Theorie und Praxis des programmierten Lernens", Darmstadt 1969, S. 75-76

intensive Laborforschung auf dem Gebiet der Verstärkung geleistet worden ist, noch die leicht beweisbare Tatsache bezweifeln, daß bemerkenswerte Verhaltensmöglichkeiten durch die geschickte Handhabung von Verstärkungsmöglichkeiten entwickelt werden können¹⁾. Ähnlich schränkt auch CROWDER die Bedeutung der Verstärkung für die Programmierete Unterweisung ein. Auf dem Hintergrund dieser Diskussion entstand eine Vielzahl von Arbeiten, in welchen die Wirkung der Verstärkung für den Lernprozeß im Rahmen der Programmierten Unterweisung untersucht worden ist²⁾. KRUMBOLTZ und WEISMAN³⁾ berichten von einer negativen linearen Beziehung zwischen der Fehlerzahl und dem Anteil der Bestätigungen an der Gesamtantwortzahl, es ergaben sich aber keine signifikanten Unterschiede zwischen den verschieden verstärkten Gruppen. BYRAN und RIGNEY⁴⁾ untersuchten drei Gruppen, von denen eine überhaupt nicht verstärkt

1) A.A. LUMSDAINE, a.a.O., S. 76

2) In diesem Zusammenhang sei, abgesehen von der bereits oben erwähnten Veröffentlichung von L.M. STOLUROW und H. LIPPERT, noch auf die Arbeiten von J. ANNET: "The Role of Knowledge of Results in Learning: A Survey", in: J.P. De CECCO (Hrsg.): "Educational Technology", New York 1964, S. 279-285 und J.D. KRUMBOLTZ und R.G. WEISMAN: "The Effect of Intermittent Conformation in Programed Instruction", in: J.P. De CECCO: "Educational Technology", New York 1964, verwiesen.

3) J.D. KRUMBOLTZ and R.G. WEISMAN, a.a.O.

4) G.L. BYRAN and J.W. RIGNEY: "An evaluation of a method of shipboard training in operations knowledge", in: J.P. De CECCO (Ed.): "Educational Technology", New York 1964, S. 594 f.

wurde, eine zweite wurde jeweils darüber informiert, ob die Lösung richtig oder falsch ist, der dritten Gruppe wurde gesagt, warum eine Antwort richtig bzw. falsch ist. Die dritte Gruppe zeigte signifikant bessere Leistungen als die beiden Vergleichsgruppen, zwischen letzteren konnten keine signifikanten Unterschiede aufgezeigt werden. Selten sind Untersuchungen, in denen die regelmäßige und sofortige Verstärkung signifikant höhere Lernleistungen erbracht hat als dies in den Vergleichsgruppen der Fall war¹⁾. ISSING stellt zusammenfassend fest²⁾, daß gerade bei dieser Gruppe von Untersuchungen nichtsignifikante Ergebnisse besonders häufig sind. Dies ist angesichts der SKINNERschen Verstärkungstheorie zweifellos ein überraschender Befund. Auch hier muß wieder festgehalten werden, daß die Befunde äußerst vielschichtig und widersprüchlich sind, obwohl gerade im Falle der Verstärkung im Hinblick auf SKINNERs Lerntheorie eindeutige Ergebnisse erwartet werden müßten. Dies kann einmal damit erklärt werden, daß die Statistik kein geeignetes Mittel ist, die Bedeutung der Verstärkung für die Verhaltensformung zu erfassen³⁾, zum anderen muß darauf hingewiesen werden, daß die widersprüchlichen Befunde durch die verschiedenen Lehrinhalte, abweichenden Adressatengruppen und divergierenden Prüfinstrumente (um nur einige Faktoren zu nennen) verursacht werden können.

-
- 1) G.W. ANGELL: "The effect of immediate knowledge of quiz results on final examination scores in freshmen chemistry", in: Journ. educ. Res., Jan. 1949, 42
 - 2) Vgl. L.J. ISSING, a.a.O., S. 108
 - 3) Vgl. B.F. SKINNER: "Fallstudie zur wissenschaftlichen Methode in der Psychologie", in: W. CORRELL (Hrsg.): "Programmiertes Lernen und Lehrmaschinen", Braunschweig 1966, S. 130 und 132

2.30 Analyse des Zusammenhangs von Lernerfolg bei Programmierter Unterweisung und Persönlichkeitsvariablen

Hierunter fallen alle Arbeiten, welche, ausgehend von Untersuchungen über die Programmierte Unterweisung, den Zusammenhang von Persönlichkeitsvariablen wie Intelligenz, Einstellungen, Angstbereitschaft usw. mit der Lernleistung untersuchen. Am häufigsten wurde in diesem Rahmen der Zusammenhang von Intelligenz und Lernleistung untersucht. Dabei ging man i. allg. von der Annahme aus, daß der Programmierte Unterricht geeignet sei, Differenzen im Lernerfolg, welche auf unterschiedlicher Schülerintelligenz beruhen, auszuschalten oder jedenfalls zu nivellieren. Hinter dieser Annahme steht die theoretische Konzeption des SKINNER-Programms und damit letztlich SKINNERS Lerntheorie. Lernleistung und Intelligenz sind für SKINNER keine benachbarten Begriffe wie dies andern Orts in der Pädagogischen Psychologie der Fall ist¹⁾, für ihn ist Lernen vielmehr gleichbedeutend mit der Wirkung einer positiven Verstärkung auf eine operante Reaktion. "Wenn wir erst einmal die(se) besondere Art einer Folgeerscheinung ... hergestellt haben, erlauben es uns unsere Methoden, das Verhalten eines Organismus fast beliebig zu formen"²⁾. Intelligenter Schüler brauchen nach SKINNER lediglich eine geringere Anzahl von Verstärkungen, sie sind daher im konventionellen Lehrerunterricht deutlich bevorzugt, da hier auf den einzelnen

1) Vgl. F. FIPPINGER: "Intelligenz und Schulleistung bei 9- und 10jährigen Volksschülern (Untersuchungsergebnisse)", in: K.-H. INGENKAMP (Hrsg.): "Schulkonflikt und Schülerhilfe", Weinheim 1965, S. 33 f.

2) B.F. SKINNER: "Die Wissenschaft vom Lernen und die Kunst des Lehrens", in: W. CORRELL (Hrsg.): "Programmiertes Lernen und Lehrmaschinen", Braunschweig 1966, S. 67

Schüler relativ wenig Verstärkungen entfallen. Da aber selbst im "klügeren Schüler ein bestimmter Verstärkungsmechanismus mehrere Male hergestellt werden muß, ehe er das Verhalten richtig in den Griff bekommen kann"¹⁾, werden darüber hinaus weniger intelligente Schüler im herkömmlichen Unterricht unter ausgesprochen schlechten Bedingungen arbeiten, da die Anzahl der auf den einzelnen Schüler entfallenden Verstärkungen selbst im Falle hoher Intelligenz nach SKINNER noch zu gering ist. Dazu kommt, daß insbesondere weniger intelligente Schüler von einem überzeugend und verständlich vorgetragenen Programm abhängig sind, das keine Lernschritte enthält, welche vom Schüler nicht bzw. nur fehlerhaft bewältigt werden können. Eben solche Programme fehlen im herkömmlichen Unterricht. "Ein ... bemerkenswerter Mangel ist das Fehlen eines durchdachten Programms, das das erstrebte komplexe Endverhalten graduell entwickelt"²⁾. Da weiter als Folge schulischen Versagens "Ängste, Unsicherheiten und Aggressionen"³⁾ auftreten, welche ihrerseits die Lernleistung negativ beeinflussen und schließlich zu einem partiellen oder totalen Resignieren des Schülers führen, müssen gerade die Lernbedingungen des weniger intelligenten Schülers im herkömmlichen Unterricht als ausgesprochen schlecht angesehen werden. Von diesem Hintergrund aus wird verständlich, daß immer wieder vermutet wurde, der Zusammenhang zwischen Intelligenz und Lernleistung im Programmierten Unterricht sei geringer als bei herkömmlichen Unterrichtsverfahren, ja sogar, daß bei Programmierter Unterweisung überhaupt

1) B.F. SKINNER, a.a.O., S. 75

2) B.F. SKINNER, a.a.O., S. 74

3) B.F. SKINNER, a.a.O., S. 74

kein signifikanter Zusammenhang erkennbar sei¹⁾. Allerdings existieren auch gegenläufige Befunde, welche den Schluß zulassen, daß die Intelligenz auch bei Programmierter Unterweisung der signifikanteste Faktor für die Aneignung von Lernmaterial ist²⁾. Widersprüchlich sind auch die Ergebnisse bei Untersuchungen, welche bei der Frage nach dem Zusammenhang von Intelligenz und Lernleistung eine vergleichende Betrachtung von Programmierter Unterweisung und konventionellem Unterricht zugrunde legen. Solche Arbeiten sind jedoch äußerst selten³⁾. In unmittelbarem Zusammenhang mit den oben ausgeführten Gedankengängen SKINNERS ist auch die häufig geäußerte Behauptung zu sehen, daß die Leistungsvariabilität bei Programmierter Unterweisung geringer ist als beim herkömmlichen Unterricht⁴⁾. Dies wurde experimentell häufig bestätigt. Zusammenfassend kann an dieser Stelle festgehalten werden, daß zwar häufig ein geringer Zusammenhang von Intelligenz und Lernleistung im Programmierten

-
- 1) Entsprechende Ergebnisse berichtet Y. FILBY: "Teaching Machines. A Review of Theory and Research", Copenhagen 1961, S. 229; L.M. STOLUROW berichtet sogar von einem negativen Zusammenhang (L.M. STOLUROW: "Teaching by Machine", Washington 1961, S. 136).
 - 2) In diesem Zusammenhang sind die Arbeiten von P. LAMBERT, D.M. MILLER and D.E. WILEY: "Experimental folklore and experimentation: The study of programmed learning in the Wauwatosa public schools", in: Journ. Educ. Res., 1962, 55, (9) und R.N. CASSEL and W.L. TELLOM: "A preliminary evaluation of programmes instruction with students of high ability", in: Psych. Rep., 1962, 10, (1), zu erwähnen.
 - 3) In diesem Zusammenhang muß auf die Arbeit von G. HUBER: "Lernpsychologische Befunde bei programmierter Unterweisung", München 1966 und K. RÖTTEL: "Vergleichende Analyse einer Versuchsreihe mit programmiertem und konventionellem Unterricht", in: ZeF 2, 1968, 4, verwiesen werden.
 - 4) Vgl. G. HUBER, a.a.O., S. 42

Unterricht festgestellt worden ist, da entsprechende Ergebnisse jedoch auch für den herkömmlichen Unterricht vorliegen, können hier nur Untersuchungen zur Klärung beitragen, welche dieser Fragestellung eine vergleichende Betrachtung von konventionellem und programmiertem Unterricht zugrunde legen. Dabei muß - jedenfalls bei Rückgriff auf die gegenwärtig verfügbaren Programme - davon ausgegangen werden, daß die Intelligenz i.allg. auch bei der programmierten Unterweisung einen signifikanten Faktor für die Aneignung von Lernstoffen darstellt.

Als weitere Persönlichkeitsvariablen wurden die Geschlechtszugehörigkeit¹⁾, die Angstbereitschaft²⁾, das Ausmaß an Extraversion und sozialer Verstärkung³⁾ und die Einstellung gegenüber der Programmierten Unterweisung⁴⁾ herangezogen. Abgesehen von Untersuchungen, welche die Variable "Einstellung" verwenden, sind solche

1) G. HUBER, a.a.O., S. 89 ff.

2) W. HOCHHEIMER: "Untersuchungen und Erprobungen zum Programmierten Lernen", in: ZeF, 1, 1967, 1

3) G.O. LEITH, M. DAVIS und N. TERENCE: "The influence of social reinforcement and personality on achievement in a school learning task", in: Research Report No. 19, National Center for Programmed Learning, Birmingham 1967

4) R.M. GAGNÉ und W. DICK: "Learning Measure in a Self-Instructional Program", in: J.P. De CECCO (Ed.): "Educational Technology, New York 1964 und G. SCHRÖTER: "Untersuchungen über die Einstellung von Berufsschülern zur Programmierten Unterweisung", in: "Praxis und Perspektiven des Programmierten Unterrichts", Quickborn 1965

Arbeiten außerordentlich selten. Diejenigen Autoren, welche den Aspekt der individuellen Lernvoraussetzungen ihrer Arbeit zugrunde legten, konzentrierten sich i.allg. auf die Variablen "Intelligenz", "Begabung" und "Lerntempo". Bezeichnenderweise erwähnt auch ISSING in seiner bibliographischen Übersicht ausdrücklich diese drei Untersuchungsbereiche. Aufgrund der geringen Anzahl einschlägiger Untersuchungen können die Probleme auf diesem Gebiet noch als völlig ungeklärt gelten. Dies gilt umso mehr, als die wenigen verfügbaren Ergebnisse nicht signifikant bzw. nicht eindeutig interpretierbar sind. Eine Ausnahme machen die Einstellungsuntersuchungen, hier zeigt sich, daß die programmierte Unterweisung i.allg. signifikant positiver beurteilt wird als der herkömmliche Vergleichsunterricht. Da dies, abgesehen von der Intelligenz, die einzige Persönlichkeitsvariable ist, welche häufig Untersuchungen zugrunde gelegt wurde, kann dieses Ergebnis als einigermaßen gesichert gelten. Dabei kann allerdings vermutet werden, daß der Hawthorne-Effekt die einschlägigen Ergebnisse grundsätzlich zugunsten der Programmierten Unterweisung verzerrt. An dieser Stelle muß, entsprechend zu den Untersuchungen, in welchen der Zusammenhang von Intelligenz und Lernleistung analysiert wurde, festgehalten werden, daß auch bei Zugrundelegung anderer Persönlichkeitsvariablen i.allg. lediglich deren Zusammenhang mit der Lernleistung im programmierten Unterricht untersucht wurde. Wir haben bereits oben darauf hingewiesen, daß dies kein geeigneter Weg ist um festzustellen, ob die Programmierte Unterweisung für bestimmte Schülergruppen optimalere Lernvoraussetzungen schafft als dies dem herkömmlichen Unterricht möglich ist, da auch innerhalb einer Unterrichtsmethode bei verschiedenen Adressatengruppen,

Lerninhalten und Prüfungsinstrumenten eine hohe Variabilität inhaltlich vergleichbarer Korrelationskoeffizienten erwartet werden muß. Die geringe Anzahl an Untersuchungen, welche Persönlichkeitsvariablen mit einbeziehen, ist einigermaßen überraschend, da die Programmierte Unterweisung mit dem Anspruch auf eine Individualisierung des Lernens aufgetreten ist und in theoretischen Abhandlungen häufig von dem Prinzip der sozialen Entlastung des Lernprozesses durch die Programmierte Unterweisung gesprochen wird. SKINNER selbst sah einen Aspekt des Versagens des herkömmlichen Unterrichts in der Technik der "aversiven Kontrolle" des Schülerverhaltens. "Eine geringfügige aversive Handlung des Lehrers ruft Reaktionen hervor, die schwere Gegenmaßnahmen erfordern und den Schüler wiederum zu heftigen Taten reizen. Diese Eskalation steigt so lange an, bis einer sich zurückzieht ...¹⁾. Weiter fährt SKINNER fort: "Weniger offen, aber darum nicht weniger wichtig als Folge der aversiven Kontrolle ist die einfache Untätigkeit" ... "alle diese Reaktionen haben emotionale Begleiterscheinungen. Angst und Furcht sind charakteristisch für Ausflucht und Meidung, Ärger für die Gegenattacke und Groll für mürrische Untätigkeit"²⁾. Ungeachtet der Tatsache, daß SKINNER hier eine Schulk Wirklichkeit schildert, die kennzeichnend für den autoritären, nicht aber den demokratischen Unterricht ist, der in unseren Schulen immer mehr praktiziert wird, bleibt doch als harter Kern dieser Aussage, daß die Lernprozesse der Schüler durch soziale Konflikte im Raum

1) B.F. SKINNER: "Warum Lehrer versagen", in: W. CORRELL (Hrsg.): "Zur Theorie und Praxis des programmierten Lernens", Darmstadt 1969, S. 95

2) B.F. SKINNER, a.a.O., S. 95

der Schule, durch die Einstellung der Schule, dem Lehrer und den Mitschülern gegenüber negativ beeinflusst werden¹⁾. Daß all dies, jedenfalls langfristig gesehen, zu einer deutlichen Beeinträchtigung der Schülerleistungen führt, wurde im Rahmen anderer Untersuchungen wiederholt bestätigt²⁾. In besonders eindringlicher Form hat CORRELL auf das Prinzip der sozialen Entlastung und der Versachlichung der Lernprozesse hingewiesen. "Die Verstärkungen werden sachlich verabreicht, wodurch wir eine Erhöhung der sachlichen und primären Motivation des Schülers erreichen. Überdies wird der Lernende vom Druck der negativen Antipathie-Beziehungen innerhalb einer Schulklasse befreit und entgeht dadurch vielen möglichen Lernstörungen"³⁾. Weiter fährt der Autor fort: "Programmiertes Lernen führt zu reiner sachlicher Arbeitshaltung, weil die Verstärkungen durchweg unpersönlich verabreicht werden. Dies ist eine wesentliche Voraussetzung für die seelische Gesundheit des arbeitenden Menschen und für seine betont sachliche Arbeitshaltung"⁴⁾. Solche Behauptungen betreffs der Vorzüge der Programmierten Unterwei-

1) Vgl. W. CORRELL: "Lernstörungen beim Schulkind", Donauwörth 1968 und K.-J. INGENKAMP (Hrsg.): "Schulkonflikt und Schülerhilfe", Weinheim 1965

2) Vgl. W. CORRELL: "Das pädagogisch-psychologische Problem der Beziehung zwischen Lehrstil und Lernleistung", in: Th. HERMANN (Hrsg.): "Psychologie der Erziehungsstile", Göttingen 1966, S. 226 ff.

3) W. CORRELL: "Zur Theorie und Praxis des programmierten Lernens", Darmstadt 1969, S. 11-12

4. W. CORRELL, a.a.O., S. 12

sung gegenüber dem herkömmlichen Unterricht können zwar theoretisch überzeugend abgeleitet werden, eine experimentelle Verifizierung bzw. Modifizierung ist bisher allerdings kaum erfolgt. Entsprechende Experimente sind umso notwendiger, als in der Schulpraxis bereits dazu übergegangen wird, im Rahmen der Unterrichtsdifferenzierung u.a. konflikthanfällige Schülergruppen bevorzugt programmiert zu unterrichten, wobei mit diesem Vorgehen die Erwartung höherer Lernleistungen verbunden wird.

2.40 Zusammenfassung

Obwohl die weitaus geringere Anzahl der in diesem Kapitel angeführten Arbeiten nicht mit dem Problem behaftet war, einen schwer zu definierenden und experimentell nur unzulänglich erfaßbaren konventionellen Vergleichsunterricht in die Untersuchung mit einzubeziehen, liegen bei fast allen Forschungsansätzen im Bereich der Programmierten Unterweisung in größerer (z.T. überwiegender) Zahl nichtsignifikante Ergebnisse vor, darüber hinaus sind die signifikanten Ergebnisse häufig keinesfalls eindeutig interpretierbar, da sie in verschiedene Richtungen weisen. Spätestens an dieser Stelle muß dem Kritiker der sog. konventionellen Vergleichsstudie deutlich werden, daß nichtsignifikante und widersprüchliche Ergebnisse keinen Grund darstellen, konventionelle Vergleichsstudien abzulehnen, es sei denn, man lehnt fast die gesamte Forschung auf dem Gebiet der Programmierten Unterweisung ab. Weiter können nichtsignifikante und widersprüchliche Befunde nicht als Erklärung dafür herange-

zogen werden, daß konventionelle Vergleichsstudien bezüglich ihrer theoretischen Fundierung unzulänglich abgesichert seien, jedenfalls solange nicht, als dieselben Kritiker für die unter 2.20 und 2.30 aufgeführten Untersuchungen dieses Argument nicht anwenden¹⁾.

Wir müssen vielmehr versuchen, für die häufig nicht-signifikanten und widersprüchlichen Befunde auf dem gesamten Gebiet der Programmierten Unterweisung Erklärungen bzw. Erklärungsansätze zu finden, um daraus für unser eigenes Vorhaben die notwendigen Folgerungen ziehen zu können.

Angesichts der widersprüchlichen experimentellen Befunde könnte der Schluß gezogen werden, daß ein experimenteller Zugriff auf dem Gebiet der Analyse von Unterrichtsmethoden und Unterrichtsmitteln heute noch nicht bzw. überhaupt nicht möglich ist. In diesem Sinne äußert sich im Zusammenhang mit einem Schulfernsehversuch HEINRICHS: "Zur Auswertung selbst darf gesagt werden, daß exakte Aussagen über die Wirkung von Unterricht nicht möglich sind. Quantitative Aussagen geben qualitativ wenig her. Paralleluntersuchungen von Unterrichtsabläufen zum gleichen Thema bei verschiedenartiger Methodenorganisation ... erweisen sich deshalb als unergiebig, weil die didaktischen Ausgangslagen der Klassen, des Führungsstils des Lehrers ... zu verschieden sind, um vergleichende Messungen ernst nehmen zu können."²⁾ Diese Aussage steht stellvertretend für jenen Teil der Pädagogik, welcher einem experimentellen Zugriff oder jedenfalls einer Quan-

-
- 1) Hier muß insbesondere auf K.J. KLAUER: "Über die Erfassung der Validität von Lehrprogrammen", in: pl, 1, 1968, S. 2, verwiesen werden.
 - 2) H. HEINRICHS: "Erste Ergebnisse des Schulfernsehversuchs Hannover", in: Deutsche Lehrprogramme für Schule und Praxis, 2, 1965, S. 48

tifizierung der experimentellen Daten negativ gegenübersteht und sich - jedenfalls z.T. - auf die traditionellen Methoden der Pädagogik zurückzieht. Demgegenüber halten wir daran fest, daß gezieltes Experimentieren im Zusammenhang mit der Erforschung der Lehr-effektivität von Unterrichtsmethoden und Unterrichtsmitteln das zentrale Verfahren darstellen muß. Die bislang unbefriedigenden Ergebnisse weisen keinesfalls darauf hin, daß sich das experimentelle Vorgehen als solches als ungeeignet erwiesen hat, vielmehr ist von dieser Grundlage auszugehen und mit Hilfe verbesserter experimenteller Ansätze und gegebenenfalls präzisieren Fragestellungen weiterzuarbeiten. In diesem Sinne äußert sich auch LUMSDAINE, einer der bekanntesten Vertreter der experimentellen Unterrichtsforschung auf dem Gebiet der Programmierten Unterweisung: "Die Sterilität, die sich oft bei gezieltem Experimentieren über Unterrichtsmittel findet, soll nicht geleugnet werden, nach meiner Meinung liegt die Schwäche jedoch nicht in erster Linie in der rationalen Grundlage des experimentellen Vergleichs an sich, sondern eher in gewissen Fehlern bei der Durchführung bestimmter Experimente"¹⁾.

Wenn wir also davon ausgehen, daß das Experiment weiter Grundlage der einschlägigen Forschungsbemühungen sein muß, und hier befinden wir uns in Übereinstimmung mit der gegenwärtigen Entwicklung, dann müssen wir versuchen, die wesentlichen grundsätzlichen Fehlerquellen eines Großteils der bisherigen Untersuchungen aufzuzeigen. Hier kann nun zunächst konstatiert werden, daß e i n e Ursache der widersprüchlichen Befunde mit Sicherheit

1) A.A. LUMSDAINE: "Probleme beim programmierten Unterricht", in: W. CORRELL (Hrsg.): "Zur Theorie und Praxis des programmierten Lernens", Darmstadt 1969, S. 86

darin zu suchen ist, daß die Untersuchungen in den wenigsten Fällen so angelegt sind, daß sie echte Vergleiche ermöglichen. Selbst partielle Wiederholungsuntersuchungen sind daher äußerst selten¹⁾. Dargebotene stoffliche Inhalte, stoffliche Strukturen, Art und Größe der Stichproben, Dauer und weitere äußere Umstände der Untersuchung, Inhalt und Struktur der Prüfungsfragen usw. weichen in nicht kontrollierbarer Art und Weise voneinander ab. Häufig werden die entsprechenden experimentellen Bedingungen nur völlig unzulänglich geschildert. Untersuchungen sollten daher so angelegt werden, daß sie möglichst partiell oder sogar total wiederholbar sind. Dazu ist es allerdings notwendig, daß die Angaben über äußere Versuchsumstände, Stichprobenzusammensetzung, unabhängige und abhängige Variablen usw. möglichst detailliert erfolgen, da wir beim heutigen Stand der Forschung davon ausgehen müssen, daß die komplexe Vielfalt der auf die Lernleistung der Schüler einwirkenden Variablen nicht einfach als unsystematische und damit zu vernachlässigende Störvarianz behandelt werden kann. Solange keine - wenigstens partielle - Wiederholungsuntersuchungen vorliegen, sollten darüber hinaus experimentelle Befunde keinesfalls in den Kontext der theoretischen Diskussion aufgenommen werden, da solche Ergebnisse irreführend sein können. Schließlich muß bei der Interpretation gesicherter Ergebnisse davon ausgegangen werden, daß Generalisierungen nur unter äußerster Zurückhaltung

1) Vgl. L.D. EIGEN: "Probleme der Forschung im Bereich des programmierten Unterrichts", in: L.J. ISSING (Hrsg.): "Der Programmierte Unterricht in den USA heute", Weinheim 1967, S. 84

versucht werden sollten. Gerade auf dem Gebiet der Unterrichtsmethodenforschung muß der Versuch, allgemeingültige Ergebnisse zu finden, zugunsten einer pragmatischen Orientierung zunächst zurückgestellt werden. Weiter stellt sich die Frage, ob die bislang verwendeten Versuchspläne und die darauf basierenden Verrechnungsmethoden den Anforderungen der behandelten Fragestellungen genügen. In diesem Zusammenhang schreibt LUMSDAINE: "... hinzu kommt das Fehlen von adäquaten Maßstäben für die Empfindlichkeit oder den Aussagewert experimenteller Vergleiche. Die zuletzt genannte Schwäche läßt die Wirkungen starker Ursachen von Variationen hinter einem Gewirr von statistisch unbedeutenden Unterschiedlichkeiten zurücktreten. Diese zahlreichen Unterschiede rühren oft von einer Vielzahl technischer Unzulänglichkeiten her, die zumeist die voreilige Durchführung gezielter Experimente ohne zureichend vorbereitende Versuche widerspiegeln. Eine der hauptsächlichen Unzulänglichkeiten liegt darin, daß man sich auf unzureichend empfindliche Instrumente für Prüfmessungen verläßt."¹⁾ An dieser Stelle spricht LUMSDAINE ein zentrales Problem der Forschung über die Programmierte Unterweisung an, nämlich die häufig nicht beachtete entscheidende Bedeutung des Tests, mit welchem die abhängige Variable "Lernleistung" gemessen wird. Diese wird daraus ersichtlich, daß selbst eine noch so gut angelegte Untersuchung mit theoretisch fundierter Fragestellung an der unzulänglichen Erfassung der abhängigen Variablen scheitern kann. Nur in wenigen Arbeiten werden differenzierte Angaben über das verwendete Prüfungsinstrument gemacht, häufig werden nicht einmal die Prü-

1) A.A. LUMSDAINE, a.a.O., S. 86

fragestellungen angegeben. Wenngleich an kriteriumsbezogene Tests nicht so ohne weiteres die Forderungen der einschlägigen Testtheorie herangetragen werden können, so muß doch eine Vielzahl von Angaben gemacht werden, um das Testinstrument ausreichend zu kennzeichnen. Wir werden im methodischen Teil unserer Arbeit noch ausführlich auf dieses Problem zurückkommen.

Schließlich kann das gegenwärtig widersprüchliche Bild der Forschungsergebnisse auf dem Gebiet der Programmierten Unterweisung damit erklärt werden, daß die Hypothesen theoretisch unzulänglich fundiert sind. Daß in diesem Fall selbst die besten experimentellen Ansätze zu keinen befriedigenden Ergebnissen führen können, ist selbstverständlich. In diesem Sinne äußert sich LUMSDAINE: "Unter diesen Fehlern fällt die geistig anspruchslöse, nicht theoretisch orientierte Art und Weise auf, in der unabhängige Variablen oft definiert oder gesetzt werden"^{1,2}). An dieser Stelle muß allerdings darauf hingewiesen werden, daß eine große Zahl der einschlägigen Arbeiten ihre Hypothesen aus Behauptungen und Konsequenzen insbesondere der Theorie der operationalen Konditionierung von SKINNER ableiten. Die Autoren können i. allg. auf eine Fülle von einschlägigen Befunden hin-

1) A.A. LUMSDAINE, a.a.O., S. 86

2) So untersuchte z.B. G. HUBER: a.a.O., S. 89 ff. die Frage, inwieweit geschlechtsspezifische Differenzen bezüglich der Lernleistung bei programmiertem und herkömmlichem Unterricht bestehen, obwohl in der einschlägigen Literatur nirgends entsprechende Hinweise zu finden sind. Der Autor unternahm auch nicht den Versuch, diese Fragestellung zu begründen.

weisen, welche ihre Hypothese sinnvoll erscheinen lassen¹⁾, d.h. eine ausreichende theoretische Fundierung muß in vielen Fällen als gegeben erachtet werden. Es ist allerdings denkbar, daß die Hypothesen, welche ja alle ausnahmslos menschliches Lernen von sinnhaltigem Material betreffen, in dem Augenblick, in welchem sie sich ausschließlich auf die Theorie der operanten Konditionierung beziehen, theoretisch nicht mehr zureichend fundiert sind, da diese Lerntheorie vorzüglich an Tieren und mit jedenfalls nicht so hochstrukturiertem, sinnhaltigem Lernmaterial gewonnen worden ist. Dahinter steht die sehr häufig diskutierte, grundsätzliche Frage nach der Übertragbarkeit tierpsychologischer Befunde auf das menschliche Lernen. Wir werden unten auf die uns interessierenden Aspekte dieser Frage noch ausführlich eingehen.

- 1) Dies gilt in allererster Linie für all diejenigen Untersuchungen, welche einen Zusammenhang zwischen der Anzahl der Verstärkungen und der Lernleistung suchten.

3.00 Bericht und methodenkritische Diskussion repräsentativer Vergleichsstudien.

Wir haben unter 2.00 lediglich in einer allgemeinen Übersicht über bisherige Forschungsschwerpunkte, deren Problematik und den Trend der einschlägigen Ergebnisse berichtet, ohne auf einzelne Arbeiten detailliert einzugehen. Anstatt in dieser allgemeinen Orientierung den Leser mit einer Vielfalt methodischer Einzelheiten und einer großen Zahl widersprüchlicher Ergebnisse zu verwirren, haben wir uns entschlossen, an dieser Stelle aus der großen Anzahl einschlägiger Arbeiten über einige ausführlich zu berichten, wobei deren Anliegen mit unserer eigenen experimentellen Fragestellung unmittelbar vergleichbar ist und die bezüglich ihrer methodischen Konzeption als repräsentativ angesehen werden können. Eine relativ ausführliche Diskussion ist notwendig, um die Details unserer eigenen Arbeit in den Kontext vergleichbarer Bemühungen zu stellen und die bisherige Praxis der Vergleichsstudien detailliert kritisch zu analysieren, um daraus die nötigen Konsequenzen für unser eigenes Vorgehen ziehen zu können. Wir werden dabei so vorgehen, daß wir zunächst alle Arbeiten sukzessiv referieren, nach diesem detaillierten informativen Überblick wollen wir kritisch kommentierend Stellung nehmen. Dieser Kritik sind, insbesondere bei den englischsprachigen Arbeiten, dort Grenzen gesetzt, wo eine Einsicht in alle Daten der jeweiligen Untersuchung notwendig wäre, wie dies z.B. bei der Frage nach der Verwendbarkeit gewisser statistischer Verrechnungstechniken der Fall ist.

3.10 G. GLEASON: "Programed learning in sixth grade arithmetic"¹⁾.

3.11 Fragestellung der Untersuchung:

Die zentrale Fragestellung galt der unterschiedlichen Lehreffektivität des herkömmlichen und des programmierten Unterrichts. Weiter wurden der Zusammenhang von Lerntempo und Lernerfolg sowie der Einfluß der Motivation und der Schulangst auf die Lernleistung bei der programmiert unterrichteten Gruppe untersucht.

3.12 Versuchsplan und Versuchsablauf:

Der Autor wählte die am häufigsten verwendete experimentelle Anordnung in der Unterrichtsforschung, indem er eine Experimental- mit einer Kontrollgruppe verglich, wobei die beiden Gruppen nach den Kriterien Intelligenz und Vorwissen "parallelisiert" wurden. Jede Gruppe umfaßte N = 164 Schüler. Beide Gruppen wurden einem Vortest unterzogen, mit welchem die unterrichtsbezogenen Vorkenntnisse gemessen wurden. Anschließend wurde die Experimentiergruppe programmiert, die Kontrollgruppe herkömmlich unterrichtet. Die Schüler der Kontrollgruppe wurden über 30 Tage hin täglich 40 Minuten unterrichtet, die Schüler der Experimentiergruppe arbeiteten ebenfalls täglich 40 Minuten, die zur Stoffbewältigung notwendige Anzahl der Tage wurde hier allerdings

1) G. GLEASON: "Programed learning in sixth grade arithmetic", Univ. of Wisconsin, Milwaukee 1964, (AERA-Paper)

den einzelnen Schülern überlassen. Nach Abschluß der experimentellen Phase wurde beiden Gruppen eine Parallelförm des Vortests vorgelegt. Nach Ablauf von 6 Wochen wurde eine weitere Parallelförm des Vortests eingesetzt um die Behaltensleistung zu messen. Die abhängige Variable "Lernleistung" wurde, wie oben schon angedeutet, mit 3 Parallelförm des eigens für diese Untersuchung entwickelten "Decimal Achievement Test" gemessen, die unabhängige Variable "Programmierter Unterricht" war durch das verwendete Linearprogramm definiert, der herkömmliche Unterricht entsprach der üblichen Form des Klassenunterrichts.

3.13 Ergebnisse der Untersuchung:

Beim Nachtest zeigte die programmiert unterrichtete Gruppe eine signifikant bessere Lernleistung, beim Behaltenstest indessen waren die Unterschiede nur noch zufällig. Hochmotivierte und wenig ängstliche Schüler der programmiert unterrichteten Gruppe lernten signifikant mehr als gering motivierte und relativ ängstliche Schüler derselben Gruppe (hier wurde das obere bzw. untere Quartil herangezogen), wobei erwähnt werden muß, daß ein hoher Zusammenhang zwischen Intelligenz und hoher Motivation bzw. geringer Schulängstlichkeit bestand (dies muß aus den vorgegebenen arithmetischen Mitteln der einzelnen Variablen innerhalb der Quartile geschlossen werden, Korrelationskoeffizienten wurden nicht angegeben). Die Interkorrelation zwischen Lernleistung und Schulängstlichkeit betrug bei der herkömmlich unterrichteten Gruppe 0.21, bei der programmiert unterrichteten Gruppe 0.19.

- 3.20 D.M. MILLER, D.E. WILIEY, P. LAMBERT: "An experimental study of the effect of instructional method on the acquisition, retention and transfer of English punctuation skills"¹⁾.

3.21 Fragestellung der Untersuchung:

Die zentrale Fragestellung galt auch hier der unterschiedlichen Lern- und Behaltensleistung bei Programmierter Unterweisung und herkömmlichem Unterricht, wobei zwischen "direktem" Lernen und der Transferleistung unterschieden wurde. Die gesamte Stichprobe wurde in drei Leistungsgruppen unterteilt um gruppenspezifische Aussagen über die Lehreffektivität der beiden Unterrichtsmethoden gewinnen zu können. Schließlich wurde noch der Einfluß der Lernbereitschaft auf die Lernleistung in die Untersuchung mit einbezogen.

3.22 Versuchsplan und Versuchsablauf:

Auch hier wurde das bereits oben geschilderte experimental design gewählt, wobei die beiden Gruppen wiederum nach Intelligenz und Vorwissen "parallelisiert" wurden. Die Untersuchung wurde im 10. (12 Klassen; N = 322) und 11. Schuljahr (12 Klassen; N = 340) durchgeführt. Der Experimentier- und der Kontrollgruppe wurden je sechs

1) D.M. MILLER, D.E. WILIEY, P. LAMBERT: "An experimental study of the effect of instructional method on the acquisition, retention and transfer of English punctuation skills", Univ. of Wisconsin, Milwaukee 1964, (AERA-Paper)

Klassen zugeteilt. Der Programmierunterricht war durch das der Untersuchung zugrunde gelegte Programm definiert, die Kontrollgruppe wurde in "Form des üblichen Klassenunterrichts" unterrichtet, wobei dieser inhaltlich streng auf das Programm hin abgestimmt war. Die Lernbereitschaft wurde dadurch differenziert, daß der einen Gruppe der Hinweis auf die Bedeutsamkeit ihrer Leistung für ihre Zensur gegeben, der anderen Gruppe indessen die gegenteilige Versicherung gegeben wurde. Die drei Leistungsgruppen wurden mittels der Kriterien Intelligenz, Schulnote und Schulleistung (Thema: Zeichensetzung in der englischen Sprache) gebildet.

3.23 Ergebnisse der Untersuchung:

Im 11. Schuljahr ergaben sich unabhängig von Leistungsgruppe und Motivation i. allg. keine signifikanten Unterschiede hinsichtlich der einfachen Lernleistung, dasselbe gilt für die Transferwirkung der beiden Unterrichtsmethoden. Lediglich in der unteren Leistungsgruppe ohne Motivation erwies sich der konventionelle Unterricht - entgegen den Erwartungen - als signifikant besser. Im 10. Schuljahr ergab sich, unabhängig von der Motivation, in der unteren Leistungsgruppe bei den programmiert unterrichteten Schülern eine signifikant höhere Lernleistung, dasselbe gilt für die mittlere Leistungsgruppe, in der oberen Leistungsgruppe waren von den motivierten Schülern die herkömmlich unterrichteten signifikant besser, in der Gruppe ohne Motivation war die Lernleistungsdifferenz zufällig. Entsprechend lagen die Werte der Transferprüfung.

3.30 A. ROE, M. MASSEY, G. WELTMAN und D. LEEDS: "Automated Teaching Methods using linear Programs"¹⁾.

3.31 Fragestellung der Untersuchung:

In dieser Arbeit wird die Auswirkung verschiedener Darbietungsmodi auf die Lehreffektivität untersucht. Unter anderem werden dabei der Unterricht mit Buchprogramm (dies fiel bislang bei uns unter die Bezeichnung "Programmierter Unterricht") und verschiedene Formen des herkömmlichen Unterrichts der Untersuchung als unabhängige Variablen zugrunde gelegt.

3.32 Versuchsplan und Versuchsablauf:

Die Untersuchung basiert auf einer Stichprobe mit 186 Ingenieurstudenten. Das durchschnittliche Alter betrug ca. 18 Jahre. Die Zuteilung zu den einzelnen Versuchsgruppen erfolgte - im Gegensatz zu den bisher referierten Untersuchungen - per Zufall (echte Randomisierung!), dabei wußten die Versuchspersonen nicht, daß sie an einem Experiment teilnahmen. Bezüglich des Unterrichtsstoffes besaßen sie keine oder nur sehr geringe Vorkenntnisse. Das verwendete Unterrichtsprogramm umfaßt 192 Lernschritte und ist linear aufgebaut. Es behandelt elementare Begriffe der Wahrscheinlichkeitsrechnung. Die beiden Formen des herkömmlichen Unterrichts wurden wie folgt defi-

1) A. ROE, M. MASSEY, G. WELTMAN und D. LEEDS: "Automated Teaching Methods using linear Programs", Report Nr. 60-105, Dezember 1960, Department of Engineering, Univ. of California, Los Angeles

niert: Zwei Gruppen wurden von jeweils einem Lehrer unterrichtet, welche beide mit Inhalt und Konstruktion des der Untersuchung zugrunde gelegten Lehrprogramms weitgehend vertraut waren und die Anweisung hatten, sich bezüglich des methodischen Vorgehens, d.h. der Sequenz der aufeinanderfolgenden Lernschritte, möglichst eng an das Programm zu halten. Es handelte sich also letztlich um eine Wiedergabe des Programms in verbaler Form. Ein dritter Lehrer unterrichtete eine weitere Gruppe. Diesem standen eine Übersicht über das Programm, ein Exemplar des Schlußtests sowie eine Reihe von Lehrbüchern, welche das behandelte Stoffgebiet enthalten, zur Verfügung. Bezüglich seines methodischen Vorgehens wurden diesem Lehrer keine Vorschriften gemacht. Hier wird also die übliche Form des Klassenunterrichts repräsentiert, wobei dieser allerdings durch die Abgrenzung zu der anderen Form des herkömmlichen Unterrichts etwas differenzierter definiert ist als dies bei den bisher berichteten Untersuchungen der Fall ist. Nach Beendigung der experimentellen Phase wurde der Lernzuwachs mit Hilfe eines 20 items (free response) umfassenden Tests gemessen.

3.33 Ergebnisse der Untersuchung:

Die experimentell ermittelten Daten wurden mittels einer Varianzanalyse verrechnet. Dabei ergaben sich keine signifikanten Unterschiede zwischen denjenigen Gruppen, die ausschließlich mit dem Buchprogramm arbeiteten und jenen Gruppen, welche von den Lehrern unterrichtet wurden, die sich streng an das Programm hielten. Dagegen ergab sich zwischen all diesen Gruppen und der Gruppe, welche in "freier Weise" herkömmlich unterrichtet wurde, jeweils ein signifikanter Unterschied auf dem 5 %-Niveau.

3.40 G. HUBER: "Lernpsychologische Befunde bei Programmierter Unterweisung"¹⁾.

3.41 Fragestellung der Untersuchung:

In dieser Untersuchung wird insbesondere die Frage nach der Auswirkung programmierten und herkömmlichen Unterrichts auf die Lern- und Behaltensleistung gestellt. Weiter wurde dem Zusammenhang von Lernleistung und Intelligenz im Rahmen der beiden Unterrichtsmethoden nachgegangen. Schließlich untersuchte der Autor noch die Frage des Zusammenhangs von Geschlechtszugehörigkeit und Lernleistung. Durchgehend wurde zwischen Schülern der beiden in den Versuch einbezogenen Schulzweige differenziert (8. Volksschulklasse und die entsprechende Schuljahrgangsstufe in der Mittelschule).

3.42 Versuchsaufbau und Versuchsablauf:

Den verrechneten Daten liegt eine Stichprobengröße von $N = 97$ zugrunde, welche sich aus vier Klassen rekrutierte. Für den Versuchsplan wurden insgesamt vier Gruppen gebildet. Davon entfielen je zwei auf die beiden Schultypen, jeweils eine davon wurde programmiert bzw. herkömmlich unterrichtet. Die einem Schultyp zugeordneten Gruppen wurden bezüglich Intelligenz und Vorwissen parallelisiert, wobei die Vergleichbarkeit entsprechend den beiden zuerst berichteten Untersuchungen dann als gegeben angesehen wurde, wenn nichtsignifikante Unterschiede bezüglich der

1) G. HUBER: "Lernpsychologische Befunde bei programmierter Unterweisung", München 1966

Kriteriumsvariablen vorlagen. Die Bearbeitungszeit der programmiert unterrichteten Gruppen war jedem einzelnen Schüler freigestellt, das arithmetische Mittel der Bearbeitungszeit betrug 128 (Volksschule) bzw. 112 (Mittelschule) Minuten, der herkömmliche Unterricht betrug 120 Minuten. Das der Untersuchung zugrunde gelegte Lehrprogramm behandelt ein Thema aus dem Bereich der Naturkunde und besteht aus 181 Lernschritten, die linear aufgebaut sind. Über den herkömmlichen Unterricht wird - abgesehen von der zeitlichen Aufteilung der Unterrichtsstunden - lediglich berichtet, daß die Lehrer sich stofflich am Unterrichtsprogramm orientieren mußten, bezüglich der weiteren methodischen Gestaltung wurden keine Auflagen gemacht. Nach Ablauf der experimentellen Phase und nach Ablauf von vier Monaten wurde die Lern- bzw. Behaltensleistung mittels eines eigens für das Programm konstruierten Tests gemessen, dieser war den Schülern auch zur Ermittlung der Vorkenntnisse vorgelegt worden. Als abhängige Variablen wurden der Prozentsatz der jeweils richtig beantworteten Testitems, der effektive Lernzuwachs in Prozenten des gesamten Lehrinhalts und der effektive Lernzuwachs in Prozenten der optimalen Lernmöglichkeit zugrunde gelegt.

3.43 Untersuchungsergebnisse:

Bei den jeweils demselben Schultyp zugehörigen Gruppen zeigte die programmiert unterrichtete in beiden Fällen signifikant bessere Lern- und Behaltensleistungen. Signifikante Unterschiede fanden sich auch zwischen den verschiedenen Schulgruppen. Die Intelligenz erwies sich beim programmierten Unterricht - im Gegensatz zum herkömmlichen - als signifikanter Faktor für die Stoffaneignung, geschlechtsspezifische Unterschiede innerhalb der beiden Unterrichtsmethoden konnten nicht nachgewiesen werden.

3.50 K. RÖTTEL: "Vergleichende Analyse einer Versuchsreihe mit programmiertem und konventionellem Unterricht"¹⁾.

3.51 Fragestellung der Untersuchung:

Verglichen wurde die Lehreffektivität von Programmierter Unterweisung und herkömmlichem Unterricht, wobei als abhängige Variablen Lern- und Behaltensleistung zugrunde gelegt wurden. Weiter untersuchte der Autor den Zusammenhang von Lernleistung und Intelligenzfaktoren, wobei er auch hier zwischen den beiden Unterrichtsmethoden differenzierte.

3.52 Versuchsaufbau und Versuchsablauf:

Der Untersuchung wurden zwei neunte Klassen eines Gymnasiums zugrunde gelegt, eine davon wurde programmiert, die andere herkömmlich unterrichtet. Unterrichtsthema war eine Einführung in die Mengenalgebra. Die Unterrichtsstunden der programmiert unterrichteten Klasse liefen wie folgt ab: Nach einer durchschnittlich 18 Minuten umfassenden Wiederholung, Zusammenfassung und Vertiefung, welche in Form des herkömmlichen Unterrichts gegeben wurde, erfolgte eine 20-minütige Stillarbeit mit einem Buchprogramm. Die herkömmliche Unterrichtsstunde umfaßte 10 Minuten gemeinsamer Wiederholung und anschließend 28 Minuten größtenteils darbietenden Lehrerunterricht. Sofern die Schüler der Programmgruppe das vorgegebene Lehr-

1) K. RÖTTEL: "Vergleichende Analyse einer Versuchsreihe mit programmiertem und konventionellem Unterricht", in: ZeF, 2, 1968, 4

ziel einer Unterrichtsstunde nicht erreichten, mußten sie dieses in der Form von Hausaufgaben nachholen, wer vor Ablauf einer Schulstunde fertig war, wurde mit vertiefenden Zusatzaufgaben versorgt. Nach jeweils einem Viertel der Untersuchungszeit wurde eine Wiederholungsprüfung angesetzt, welche in der anschließenden Stunde besprochen wurde. Drei Monate nach Abschluß der experimentellen Phase wurde ein Behaltenstest durchgeführt. Abgesehen von einer Aufzählung der geprüften Themenkreise liegen über die Tests keine weiteren Angaben vor. Die Schüler hatten bezüglich des unterrichteten Stoffes keine einschlägigen Vorkenntnisse. Das der Untersuchung zugrunde gelegte Programm ist linear aufgebaut.

3.53 Untersuchungsergebnisse:

Die Verrechnung erfolgte z.T. über Prozentwerte und "Verhältniszahlen" der richtig beantworteten Prüfungsfragen, statistische Prüfverfahren wurden nicht angewandt. Eine Analyse der solchermaßen dargestellten Daten ergibt, daß weder bei der Lern- noch bei der Behaltensleistung signifikante Differenzen vorliegen, die Lernleistungsstreuung bei der konventionell unterrichteten Gruppe ist jedoch (insbesondere im unteren Leistungsbe- reich) deutlich größer. Beim Behaltenstest fiel die programmiert unterrichtete Gruppe stärker ab, bei den Transferfragen erbrachten die herkömmlich unterrichteten Schüler eine bessere Leistung (die Richtigwertsummen verhalten sich hier wie 1,42:1; der Autor arbeitet, wie wir schon oben angedeutet haben, durchgängig mit solchen Quotienten). Die Korrelationen zwischen Intelligenz und Lernleistung waren in der konventionell unterrichteten Gruppe etwas höher (um 0.50) als in der programmiert

unterrichteten Gruppe, in beiden Fällen muß - auch unter Berücksichtigung der relativ kleinen Stichproben - auf signifikanten Zusammenhängen ausgegangen werden.

- .60 L.J. ISSING und C. SCHELLENBERG: "Anwendung von PU-Prinzipien auf die Gestaltung von Schulfernsehsendungen - eine Vergleichsuntersuchung"¹⁾.

1.61 Fragestellung der Untersuchung:

Untersucht wurde die Auswirkung von drei Darbietungsmodi (Lehrprogrammgruppe, programmierte Schulfunksendung und konventionell aufgebaute Schulfunksendung) auf die Lern- und Behaltensleistung der Versuchsgruppen.

3.62 Untersuchungsaufbau und Untersuchungsablauf:

Der Untersuchung wurden 2 Volksschulklassen des 6. Schülerjahrgangs zugrunde gelegt. Die Schüler wurden nach dem Zufallsprinzip in 3 Untersuchungsgruppen zugeteilt. Die Dauer der experimentellen Phase betrug 40 Minuten, der Fragebogen für die Prüfung des Vorwissens, Lernerfolgs und Behaltens (nach 4 Wochen) umfaßte 26 Items, diese wurden einer Itemanalyse unterzogen, Angaben über die innere Konsistenz des Tests liegen vor. Der Test wurde speziell für diese Untersuchung entwickelt. Alle drei Unterrichtseinheiten lehnten sich eng an ein Buchprogramm mit dem Thema "Magnetismus" an; dieses Programm

1) L.J. ISSING und C. SCHELLENBERG: "Anwendung von PU-Prinzipien auf die Gestaltung von Schulfernsehsendungen - eine Vergleichsuntersuchung", in: pl, 2, 1969

wurde für das Experiment auf 33 Lernschritte zusammengestrichen, es ist linear und nach der Konstruktionsantwortmethode aufgebaut. Eine der drei Gruppen arbeitete mit dem Buchprogramm, die programmierte Schulfernsehsendung wurde in engster Anlehnung an das Programm entworfen. Auch hier mußten die Schüler die Antworten aktiv formulieren und bekamen danach die richtige Lösung geboten, im Gegensatz zum Buchprogramm war hier das Lerntempo allerdings extern gesteuert. Die konventionelle Schulfernsehsendung entsprach bezüglich Aufbau, Inhalt und Text der programmierten Sendung, sie war allerdings rein darbietend, die Tafel mit den Lerneinheiten und den darin enthaltenen Aufforderungen zum aktiven Antworten wurde nicht gezeigt.

3.63 Ergebnisse der Untersuchung:

Die Verrechnung erfolgte auf der Basis der Punktwertdifferenzen zwischen Vor- und Nach- bzw. Behaltenstest mit Hilfe des MANN-WITNEY-U-Tests. Dabei ergaben sich bei der Wissensprüfung zwischen Buchprogrammgruppe und programmierter Schulfernsehgruppe keine signifikanten Unterschiede, dagegen war der Unterschied zwischen diesen Gruppen und der konventionellen Schulfernsehgruppe jeweils signifikant auf dem 1 %-Niveau, dasselbe Ergebnis zeigte sich auch bei den Werten des Behaltenstests.

3.70 Diskussion der berichteten Untersuchungen:

Wir werden im folgenden, wie wir oben bereits ausgeführt haben, zu wesentlichen methodischen Aspekten kritisch Stellung nehmen. Ergebnisse werden nur vereinzelt in die Diskussion einbezogen. An dieser Stelle muß ausdrücklich darauf hingewiesen werden, daß die von uns oben berichteten Arbeiten nicht bezüglich ihrer Ergebnisse sondern ausschließlich bezüglich des experimentellen Vorgehens repräsentativ sind. Daß die in unserem Fall deutlich überwiegend zugunsten der Programmier-ten Unterweisung ausfallenden Ergebnisse nicht repräsentativ sind, geht bereits aus unserer allgemeinen Übersicht (vgl. 2.10) hervor.

3.71 Diskussion der Versuchspläne:

Allen Untersuchungen wurde eine einfache Kontrollgruppenanordnung oder eine Variation derselben zugrunde gelegt. In jedem Fall wurde jede unabhängige Variable lediglich durch e i n e Teilstichprobe repräsentiert. Methodologisch ist gegen dieses Verfahren nichts einzuwenden, solange eine echte Randomisierung vorliegt. Dies ist jedoch nur bei den Arbeiten von ROE u.a. und ISSING und SCHELLENBERG der Fall. In allen anderen Fällen handelt es sich um eine quasiexperimentelle Anordnung, genauer gesagt um eine "Anordnung mit nicht gleichartiger Kontrollgruppe"¹⁾. Trotz der in den meisten Fäl-

1) E. SCHWARZ: "Experimentelle und quasi-experimentelle Anordnungen in der Unterrichtsforschung", in: K.-H. INGENKAMP (Hrsg.): "Handbuch der Unterrichtsforschung Teil I", Weinheim-Berlin-Basel, 1970, S. 562 ff.

len vorgenommenen "Parallelisierung" werden sich bei Wahl dieser quasi experimentellen Anordnung die Nachtestergebnisse völlig unabhängig von den experimentellen Einflüssen unterscheiden. Dies gilt umso mehr, als zur Parallelisierung nicht die in der Statistik üblichen Techniken herangezogen wurden¹⁾, sondern die untersuchten Gruppen lediglich daraufhin überprüft wurden, ob sich die arithmetischen Mittel bzw. die Streuungen der Kriteriumswerte "Intelligenz" und "Vorwissen" signifikant unterscheiden. Im Falle nichtsignifikanter Unterschiede wurden die Gruppen als "parallelisiert" betrachtet. "Der Unterschied wird proportional zu dem Unterschied zwischen den Populationen, aus denen die Auswahl erfolgte, variieren ..." ²⁾. Wir werden auf dieses Problem weiter unten noch sehr ausführlich eingehen. An dieser Stelle soll nur vermerkt werden, daß diese quasiexperimentelle Anordnung offensichtlich mit dem echten Kontrollgruppenexperiment, welchem eine Randomisierung zugrundeliegt, gleichgesetzt wird. Dies muß daraus geschlossen werden, daß weder bei der Begründung eines solchen Versuchsplans - soweit eine solche überhaupt erfolgte - noch bei der Diskussion der Ergebnisse ein Hinweis auf die Grenzen der Aussagekraft einer solchen Versuchsanordnung erfolgte. Weiter kann die stillschweigende Gleichsetzung daraus geschlossen werden, daß nicht diejenigen statistischen Verrechnungstechniken angewandt werden, welche wesentliche Fehlerquellen dieser Versuchsanordnung hätten teilweise aus-

1) Vgl. E. MITTENBECKER: "Planung und statistische Auswertung von Experimenten", Wien 1963, S. 29 f. und 115 ff.

2) E. SCHWARZ, a.a.O., S. 567

schließen können¹⁾. Schließlich hätte bei Nichtgleichsetzung des echten und des quasiexperimentellen Kontrollgruppenexperiments erwartet werden können, daß auf die häufige Voraussagbarkeit der Fehlerrichtung beim quasiexperimentellen design im Rahmen der Ergebnisinterpretation eingegangen wird. All dies ist nicht der Fall, die Autoren gehen offensichtlich davon aus, daß eine "Parallelisierung" nach den Kriterien Intelligenz und Vorwissen eine echte Randomisierung ersetzen könne. Dies ist eine falsche Grundannahme, welche für eine Vielzahl von Experimenten im Rahmen der Unterrichtsforschung kennzeichnend ist²⁾. Damit soll nicht die Verwendbarkeit dieser Form einer quasiexperimentellen Anordnung bestritten werden, es soll lediglich darauf hingewiesen werden, daß bei Zugrundelegung einer solchen dies wenigstens im Rahmen der Interpretation berücksichtigt und bei Besprechung der methodischen Probleme der Arbeit erwähnt werden müßte. Selbst im Falle einer echten Kontrollgruppenanordnung sind "Wiederholungen und Kreuzvalidierungen zu anderen Zeitpunkten und unter anderen Bedingungen erforderlich"³⁾, bevor die Ergebnisse vielschichtiger, hochstrukturierter Untersuchungen - und um diese handelt es sich angesichts der hochstrukturierten und schwer definierbaren unabhängigen und abhängigen Variablen - in den Kontext der theoretischen Diskussion aufgenommen werden können. Dies gilt nun für quasiexperimentelle Anordnungen in ganz besonderem Maße, da der Einfluß der Stichprobenverzerrung und der daraus resultierenden Wechselwirkungen

1) Vgl. E. SCHWARZ, a.a.O., S. 567

2) Vgl. E. SCHWARZ, a.a.O., S. 451 f.

3) E. SCHWARZ, a.a.O., S. 455

auf die innere Validität einer Untersuchung nicht kontrolliert ist und diese in jedem Fall mehr oder weniger stark vermindert. Wir meinen, daß im Zusammenhang mit dieser Problematik sich für die einschlägige Forschung die Aufgabe stellt, neue experimentelle bzw. quasiexperimentelle Versuchspläne zu entwickeln bzw. anzuwenden, die bezüglich der inneren Validität bessere Voraussetzungen bieten. Daß quasiexperimentelle Anordnungen trotz ihrer geringeren Exaktheit vorgezogen werden können, ergibt sich im Zusammenhang mit der äußeren Validität einer Untersuchung. Der Verzicht auf natürliche Klasseneinheiten zugunsten von randomisierten Stichproben kann vor allem bei der Variable "herkömmlicher Unterricht" zu einer starken Verminderung der äußeren Validität führen, da der herkömmliche Unterricht mittels einer Zufallsstichprobe nicht mit dem Unterricht in einer natürlichen Klasseneinheit verglichen werden kann. Besonders bedeutsam ist dabei im Falle des Rückgriffs auf Zufallsstichproben, daß die Richtung des Fehlers kaum vorausgesagt werden kann. Allerdings kann weder ROE u.a. noch ISSING und SCHELLENBERG der Vorwurf gemacht werden, sie hätten zugunsten einer echten Randomisierung die "Echtheit der pädagogischen Situation" aufgegeben, da im ersten Fall gar keine natürliche Klasseneinheit bzw. kein entsprechender Verband vorhanden war und im zweiten Fall davon ausgegangen werden kann, daß bei einer Schulfernsehsendung kein Lehrer-Schüler-Gespräch bzw. unterrichtsbezogene Interaktionen innerhalb des Klassenverbandes stattzufinden pflegen. Bei Zugrundelegung eines herkömmlichen Unterrichts hat dagegen die Randomisierung den gravierenden Nachteil, daß eine Übertragbarkeit der solchermaßen ermittelten Ergebnisse auf die "Schulwirklichkeit" nur noch sehr beschränkt - wenn überhaupt - möglich ist¹⁾.

1) Vgl. G. DIETRICH: "Bildungswirkungen des Gruppenunterrichts", München 1969, S. 52 ff.

Angesichts dieses Sachverhalts werden in der Unterrichtsforschung weitaus überwiegend quasie xperimentelle Anordnungen verwendet.

Bezüglich der Stichprobengröße können keine einheitlichen Vorschriften aufgestellt werden, da diese von der Fragestellung und der verwendeten statistischen Verrechnungstechnik abhängt. Sicher scheint uns allerdings zu sein, daß der Rückgriff auf nur eine Klasse (bei gleichzeitiger Verwendung einer quasiexperimentellen Anordnung) pro unabhängiger Variable, wie dies bei K. RÖTTEL der Fall ist, vermieden werden sollte. Dies gilt insbesondere dann, wenn Korrelationen errechnet und diskutiert werden.

Die Frage, ob die Zeit der programmiert arbeitenden Gruppe extern gesteuert, d.h. experimentell in Übereinstimmung mit dem für die herkömmlich unterrichtete Gruppe benötigten Zeitaufwand festgelegt werden soll, wird verschieden beantwortet. Sicher ist, daß eine teilweise Verlagerung der Untersuchung in den häuslichen Bereich, wie dies bei RÖTTEL der Fall ist, abgelehnt werden muß, da in diesem Falle keinerlei Kontrolle des zeitlichen Arbeitsaufwands mehr möglich ist. Das Vorgehen von HUBER und GLEASON wird dagegen in der einschlägigen Forschung häufig praktiziert, wobei man sich dabei i. allg. darauf beruft, daß das intern gesteuerte Lerntempo ja geradezu kennzeichnend für die Programmierte Unterweisung sei. Angesichts des ohne Zweifel vorhandenen Zusammenhangs von Arbeitszeit und erzielter Leistung scheint es uns jedoch angemessener zu sein, den Schülern zwar ihr individuelles Lerntempo zu belassen, ihnen jedoch zur Bewältigung des Stoffes eine obere Grenze zu setzen, welche mit dem Zeitaufwand für den herkömmlichen Unterricht genau überein-

stimmt. Die schneller arbeitenden Schüler der Programmgruppe können in der verbliebenen Zeit wiederholen. Dies entspricht auch - im Gegensatz zu dem Vorgehen von HUBER und GLEASON - den Realitäten und Bedürfnissen der gegenwärtigen Schulorganisation¹⁾. Da gerade im Rahmen der einschlägigen Unterrichtsforschung die Zahl der Störvariablen sehr groß ist, sollte man, nach unserem Dafürhalten, wenigstens ein so elementares Kriterium, wie es der zeitliche Arbeitsaufwand darstellt, vergleichbar halten. Wenn man trotzdem das Vorgehen von HUBER und GLEASON wählt, muß dies bei der Interpretation der Ergebnisse unbedingt berücksichtigt werden. Im Zusammenhang mit dem Zeitfaktor steht auch die Frage, über welchen Zeitraum hin eine Untersuchung angelegt werden sollte. Je kürzer der Zeitraum ist, desto geringer sind die zwischenzeitlich bedingten Störvariablen anzusehen, gerade bei langfristig angelegten Untersuchungen können diesbezüglich erhebliche Fehlerquellen auftreten. Umgekehrt besteht bei Wahl eines kurzen Zeitraums die Gefahr, daß der Novitätseffekt zu einem nicht mehr kontrollierbaren Ausmaß an Fehlervarianz führt, wobei auch hier - streng genommen - nicht einmal die Richtung des Fehlers mit Sicherheit vorausgesagt werden kann. Gerade bei Unterrichtsmethoden betreffenden Experimenten, bei denen eine bekannte mit einer völlig oder weithin unbekannten Methode verglichen wird, sollte die Wichtigkeit des Novitätseffekts nicht unterschätzt werden²⁾.

-
- 1) Vgl. R. GLASER, J.H. REYNOLDS und M.G. FULLICK: "Programmierter Unterricht in der Klasse", in: W. CORRELL (Hrsg.): "Programmiertes Lernen und Lehrmaschinen", Braunschweig 1966, S. 286
 - 2) Vgl. K.J. KLAUER: "Über die Erfassung der Validität von Lehrprogrammen", in: pl, 1, 1968, S. 2

Eine Untersuchungsdauer von nur 40 Minuten, wie dies bei ISSING und SCHELLENBERG der Fall ist, scheint uns auf alle Fälle zu kurz zu sein, diesbezüglich vergleichbare Untersuchungen finden sich auch äußerst selten. Auch der Zeitaufwand von HUBER muß noch als relativ gering veranschlagt werden. Umgekehrt scheint uns bei dem Zeitaufwand von GLEASON und RÖTTEL die Gefahr von zwischenzeitlichen Störvariablen recht groß zu sein, noch viel mehr gilt dies für langfristig angelegte Untersuchungen, wie sie von GLASER, REYNOLDS und FULLICK¹⁾ und SCHRAMM²⁾ berichtet werden. Wir vertreten dabei die Ansicht, daß bei solchermaßen langfristig angelegten Studien mittels zusätzlicher Kontrollgruppen der Einfluß zwischenzeitlicher Ereignisse und der Reifung³⁾ kontrolliert werden sollte.

Schließlich sei noch auf das Vorgehen von ROE u.a. hingewiesen, bei welchen die Versuchspersonen nicht wußten, daß sie an einem Experiment teilnahmen. Dies ist zweifellos ein optimaler Ansatz, der allerdings im Rahmen der Unterrichtsforschung nur selten verwirklicht worden ist und verwirklicht werden kann⁴⁾. Dazu müßte vorausgesetzt sein, daß bei der Untersuchung nur der Lehrer in Erscheinung tritt, was aus versuchstechnischen

-
- 1) R. GLASER, J.H. REYNOLDS und M.G. FULLICK; a.a.O., S. 285-321
 - 2) W. SCHRAMM (Hrsg.): "Four Case Studies of Programed Instruction", New York 1964
 - 3) Vgl. E. SCHWARZ, a.a.O., S. 460
 - 4) Eine in dieser Hinsicht besonders interessante Arbeit stellt die Veröffentlichung von E.B. PAGE: "Teacher comments and student performance. A seventy-four classroom experiment in school motivation", in: Journ. educ. Psychol., 1958, 48, S. 173-181, dar.

Gründen häufig nicht möglich ist. Es ist kein Zufall, daß gerade die Arbeit von ROE u.a. nicht im Rahmen der Schule durchgeführt worden ist. In sämtlichen anderen Fällen war den Schülern bekannt, daß sie an einem Experiment teilnahmen, wobei die Vorkehrungen, trotzdem eine künstliche experimentelle Situation zu vermeiden, verschieden umfangreich und erfolgreich waren.

3.72 Definition der unabhängigen Variable "herkömmlicher Unterricht":

Hier fällt, abgesehen von den Arbeiten von ROE u.a. und ISSING und SCHELLENBERG auf, daß der Definition des herkömmlichen Vergleichsunterrichts äußerst wenig Raum gewidmet wird. Dies ist für den einschlägigen Forschungsbereich kennzeichnend. Der herkömmliche Unterricht wird meistens als "die übliche Form des Klassenunterrichts" (HUBER, GLEASON, MILLER u.a.) bezeichnet, also als eine von Schule zu Schule und insbesondere von Land zu Land variierende "Mischung" von dozierendem Lehrergespräch, erarbeitendem Lehrgespräch, Diskussion zwischen Lehrer und Schülern, Gruppenarbeit u.a. Die Lehrer erhalten dabei lediglich die Auflage, sich inhaltlich am Programm zu orientieren, d.h. denselben Stoff zu unterrichten, wie sie dabei verfahren, bleibt ihnen überlassen. Durch dieses Vorgehen wird bei Vergleichsstudien i. allg. die programmiert unterrichtete Gruppe umso mehr begünstigt, je unabhängiger der herkömmlich unterrichtende Lehrer vom Programm arbeitet, da er den zur Erreichung der Lernziele (insbesondere der Teillernziele der einzelnen Lernschritte) notwendigen Stoff nur noch "ungefähr" vermittelt. Wenn

dann bei der Nachprüfung, welche fast immer direkt auf das Programm hin angelegt ist und häufig sogar bis in die Formulierung dem Programminhalt gleicht, die programmiert unterrichtete Gruppe signifikant besser abschneidet, heißt das zunächst nicht, daß die herkömmlich unterrichtete Vergleichsgruppe weniger gelernt hat, vielmehr muß mit hoher Wahrscheinlichkeit angenommen werden, daß sie etwas "anderes" gelernt hat. Es ist selbstverständlich, daß diese beiden Varianzanteile im Nachhinein nicht mehr aufgegliedert werden können. Eine sinnvolle Vergleichsstudie setzt daher voraus, daß bezüglich des inhaltlichen Aspekts des Unterrichts die herkömmlich unterrichtenden Lehrer sich streng an das Programm halten. Das ist bei Globalinstruktionen wie "Sie müssen sich inhaltlich an das Programm halten" keinesfalls gewährleistet, vielmehr muß hier das übergeordnete Lernziel in eine große Zahl von untergeordneten Lernzielen, welche kleinste Unterrichtseinheiten betreffen, unterteilt und dem herkömmlich unterrichtenden Lehrer als Unterrichtsgrundlage gegeben werden. Der Verzicht auf eine Definition des herkömmlichen Unterrichts muß bei Untersuchungen, welche Unterrichtsmethoden bezüglich ihrer Lehreffektivität vergleichen, weiter im Hinblick darauf kritisiert werden, daß - gerade in diesem Problemzusammenhang - hinlänglich bekannt sein müßte, daß es auch innerhalb "der üblichen Form des Klassenunterrichts" deutlich voneinander abhebbare Methoden gibt¹⁾, die sich auch bezüglich ihrer-Lehreffizienz deutlich unterscheiden²⁾. Es ist daher durchaus von Bedeutung,

1) Vgl. K. STÖCKER: "Neuzeitliche Unterrichtsgestaltung", München 1970, S. 206 ff.

2) An dieser Stelle sei auf die Veröffentlichung von G. DIETRICH a.a.O. hingewiesen.

welche Form des herkömmlichen Unterrichts zum Vergleich herangezogen wird. An dieser Stelle soll die Schwierigkeit, den herkömmlichen Unterricht im Sinne einer experimentell gesetzten unabhängigen Variable zu definieren, nicht unerwähnt bleiben. Diese Schwierigkeit enthebt aber die Autoren von Vergleichsstudien keinesfalls der Aufgabe, den herkömmlichen Unterricht wenigstens soweit zu definieren, als es die gegenwärtigen unterrichtswissenschaftlichen Möglichkeiten zulassen. Wir werden auf die einschlägigen Probleme unten noch ausführlich zurückkommen. Im Zusammenhang mit der hier stattfindenden methodenkritischen Diskussion muß ganz klar gesagt werden, daß die Untersuchungen von HUBER, GLEASON und MILLER sich den Vorwurf gefallen lassen müssen, daß weitgehend unbekannt ist, mit was die Autoren nun eigentlich ihren Programmierten Unterricht vergleichen. Der Rückgriff auf das Argument, daß man eben von einem "durchschnittlichen" herkömmlichen Unterricht ausgehen müsse, muß abgelehnt werden, da die vorhandenen, durchaus eindeutig voneinander abhebbaren herkömmlichen Unterrichtsformen bezüglich ihrer Lehreffektivität nicht einfach "gemittelt" werden können, weil sie sowohl im Hinblick auf die Formen der Lehrer- und Schüleraktivität (insbesondere deren quantitative Ausprägungen) als auch der Unterrichtssteuerung nicht ohne weiteres vergleichbar sind. An dieser Stelle sei nur vermerkt, daß das gruppenunterrichtliche Verfahren bezüglich der Steuerung der Schüleraktivität mit dem vom Lehrer geführten Unterrichtsgespräch weniger Gemeinsamkeiten hat als letzteres mit der Programmierten Unterweisung. Der Rückgriff auf einen "durchschnittlichen herkömmlichen Unterricht" ist nur dann - und selbst hier nur bedingt - erlaubt, wenn die "Unterrichte" bezüglich

unterrichtsrelevanter Kriterien vergleichbar sind und wenn möglichst viele Klassen und damit möglichst viele "Unterrichte" den Untersuchungsergebnissen zugrunde liegen. Besonders problematisch ist das Vorgehen von RÖTTEL, da hier Programmierte Unterweisung, Lehrervortrag, Lehrer-Schüler-Gespräch und Hausaufgaben "gemischt" werden. Hier kann im Rückblick niemand mehr sagen, welchen Anteil an der Lernleistung die einzelnen unterrichtlichen Aspekte bewirken. Schließlich muß im Zusammenhang mit der mangelhaften Definition der unabhängigen Variable "herkömmlicher Unterricht" noch darauf hingewiesen werden, daß unter dieser Voraussetzung sowohl die theoretische Fundierung der Hypothesen als auch die Diskussion der Ergebnisse nur unzulänglich erfolgen kann. Dazu wäre es nämlich notwendig, daß diejenigen Aspekte der unabhängigen Variablen deutlich gekennzeichnet würden, von welchen man eine direkte Auswirkung auf die abhängige Variable erwartete.

3.73 Definition der abhängigen Variablen:

Sämtliche Untersuchungen verwendeten zur Erfassung der abhängigen Variable "Lernleistung" eigens für die Untersuchung bzw. das ihr zugrunde gelegte Programm konstruierte Tests. Abgesehen von der Arbeit von ISSING und SCHELLENBERG liegen über Schwierigkeitsindex und Trennschärfe der items sowie die Reliabilität der Gesamttests keine Angaben vor. Validitätskoeffizienten werden auch von ISSING und SCHELLENBERG nicht angegeben. Angesichts der zentralen Bedeutung des Testinstruments zur Erfassung der abhängigen Variable für das Ergebnis der Untersuchung muß hier die Frage aufgeworfen werden, ob der weitgehende Verzicht auf die Prinzipien der Testkonstruk-

tion gerechtfertigt ist. Tatsächlich ist dieser Verzicht in den einschlägigen Bereichen der Unterrichtsforschung weit verbreitet. Dies wird einmal damit begründet, daß die schematischen Selektionstechniken der Testkonstruktion dazu führen, daß das Abzielen auf einen hohen Trennschärfeindex zu einer Homogenisierung der items führt, wobei Aufgaben, die höhere kognitive Anforderungen stellen, eliminiert werden. Dasselbe gilt auch für Aufgaben, die mechanisches Wissen prüfen und demzufolge nur einen sehr niedrigen Schwierigkeitsgrad aufweisen. Demgegenüber wird darauf verwiesen, daß items, welche gewünschte Lernziele repräsentieren, unbesehen vom Schwierigkeitsgrad und der statistisch ermittelten Trennschärfe in den Programmtest mit aufgenommen werden sollen, da die durch sie repräsentierten Lernziele für die Qualifikation des Tests wichtiger seien als testtheoretisch optimale Eigenschaften. Angesichts der erfolgreichen Bemühungen von BLOOM¹⁾ um die hierarchische Gliederung von Lernzielen und den Ausführungen von MAGER²⁾ über deren operationale Formulierung wurde im Verlauf einer internationalen Arbeitstagung in Berlin gefordert, "bei Schulleistungstests auf eine empirische Aufgabenanalyse zu verzichten und sich mit Expertenurteilen über die inhaltliche Angemessenheit der Aufgaben zu begnügen"³⁾. Wir halten diese Forderung für die normbezogenen Schulleistungstests für ausgesprochen einseitig, im Zusammenhang mit u n s e r e m Problem muß allerdings darauf hingewiesen werden, daß wir es nicht mit norm- sondern mit kriteriumsbezogenen Tests

1) B.S. BLOOM (Hrsg.): "Taxonomy of educational objectives", New York 1956

2) R.F. MAGER: "Preparing Objectives for Programmed Instruction", San Francisco 1962

3) K.-H. INGENKAMP: "Das Testen kognitiver Fähigkeiten und Leistungen", in K.-H. INGENKAMP: "Handbuch der Unterrichtsforschung Teil I", Weinheim-Berlin-Basel 1970, S. 1009

zu tun haben. Darüber, daß für die kriteriumsbezogenen Tests die bisherigen Methoden der Aufgabenanalyse nur von sehr beschränktem Wert sind, herrscht weitgehend Einigkeit¹⁾. Wir werden auf diese Probleme in unserem methodischen Teil noch ausführlich zu sprechen kommen. An dieser Stelle muß nur in Anschluß an das oben Ausgeführte vermerkt werden, daß der in unseren referierten Arbeiten fast durchgängig festgestellte Verzicht auf die Anwendung der Prinzipien der Testkonstruktion wissenschaftlich durchaus gerechtfertigt werden kann. Bei Rückgriff auf das Expertenurteil bei der Auswahl der items müßte diese allerdings von mehreren Experten vorgenommen werden und darüber hinaus sollten weitere methodisch abgesicherte Kontrollvorkehrungen betreffs der optimalen item-Auswahl entwickelt werden. Insbesondere das Vorgehen von RÖTTEL erscheint uns hier sehr kritikbedürftig zu sein. Angesichts des oben erwähnten Verzichts auf die Prinzipien der Testkonstruktion kann auch die Verwendung von relativ wenigen items (i.allg. bewegt sich die Anzahl im Bereich von 15 bis 30) gerechtfertigt werden. Sicher scheint jedoch zu sein, daß auch beim kriteriumsbezogenen Test möglichst viele items herangezogen werden sollten, da mit wachsender Anzahl die Teillernziele des Lehrprogramms immer vollständiger repräsentiert werden. Dagegen muß bei allen von uns referierten Arbeiten kritisiert werden, daß sich die Tests ausschließlich am Lehrprogramm orientieren. Dies liegt wohl daran, daß solche Tests bereits im Rahmen der Programmkonstruktion entwickelt werden. Diese starke Programmorientierung geht z.T. bis in die Formulierung der einzelnen items hinein. Insbesondere bei Vergleichsuntersuchungen wird durch dieses Vorgehen, wie wir schon oben

1) Vgl. K.-H. INGENKAMP, a.a.O., S. 1025 f.

angedeutet haben, die Programmgruppe begünstigt und somit eine systematische Fehlervarianz bewirkt, welche im Rahmen der einschlägigen Arbeiten i.allg. die Wahrscheinlichkeit, die Nullhypothese verwerfen zu müssen, erhöht. Demgegenüber meinen wir uns der Forderung von LUMSDAINE anschließen zu müssen, der eine Unabhängigkeit von "Programm"- und "Test"-items für unbedingt notwendig hält¹⁾. Voraussetzung ist hier allerdings, daß die einzelnen Lehrzieldimensionen eindeutig festgelegt und operational formulierbar sind, Bedingungen, welche bei einem Lehrprogramm i.allg. erfüllt sind. Weiter ist kritisch zu vermerken, daß die wenigsten Arbeiten bei der Erfassung der abhängigen Variable "Lernleistung" zwischen mechanischem Wissen und Transferleistung unterscheiden, bei den von uns berichteten Arbeiten ist dies nur bei MILLER u.a. und RÖTTEL der Fall. Dies scheint uns insbesondere bei Vergleichsuntersuchungen von besonderer Bedeutung zu sein, da seitens der Vertreter der herkömmlichen Methode der Programmierten Unterweisung häufig vorgeworfen wird, sie vermittele nur mechanisches Wissen. Um sich mit diesem Vorwurf auseinanderzusetzen bedarf es einer entsprechend differenzierten Leistungsprüfung, wobei natürlich knappe Hinweise, wie sie bei RÖTTEL zu finden sind²⁾, nicht genügen. Vorbildlich in dieser Hinsicht ist die im deutschsprachigen Raum vor kurzem veröffentlichte Arbeit von STRITTMATTER³⁾. Abschließend sei im Zusammenhang mit dem Testinstrumentarium noch vermerkt, daß angesichts der zentralen Bedeutung des Meßinstruments gefordert wer-

-
- 1) Vgl. A.A. LUMSDAINE: "Assesing the effectiveness of instructional programs", in: R. GLASER (Ed.): "Teaching Machines and programed learning, II", New York 1965, S. 296
 - 2) Vgl. K. RÖTTEL, a.a.O., S. 234
 - 3) Vgl. P. STRITTMATTER, a.a.O., S. 101 ff. und 157 ff.

den muß, daß in jeder einschlägigen Veröffentlichung möglichst detaillierte Angaben über Konstruktion und Inhalt des der Untersuchung zugrunde gelegten Testinstruments mitgeteilt werden, eine Bedingung, welche die wenigsten Arbeiten erfüllen. Dies liegt z.T. wohl daran, daß die Frage des Meßinstruments in der einschlägigen Diskussion bislang vernachlässigt worden ist¹⁾. Vorbildlich sind in dieser Hinsicht die Untersuchungen von HUBER und ISSING und SCHELLENBERG.

3.74 Statistische Verrechnungstechniken:

Einer methodenkritischen Diskussion sind auf diesem Gebiet deutliche Grenzen gesetzt, weil dazu - abgesehen von groben Fehlern - ein detaillierter Einblick in alle Untersuchungsunterlagen notwendig wäre. Grundsätzlich kann allerdings vermerkt werden, daß häufig ziemlich unreflektiert parametrische Verfahren herangezogen werden, ohne daß dabei geprüft wird, ob die einschlägigen Voraussetzungen erfüllt sind²⁾. Im Zusammenhang mit den von uns referierten Arbeiten kann darauf hingewiesen werden, daß der Rückgriff auf Verhältniszahlen, wie dies bei RÖTTEL sehr häufig der Fall ist, nicht den Anforderungen genügt, die man an wissenschaftlich fundierte Ergebnisse stellen muß. Auch die von HUBER praktizierte Technik, Differenzwerte und Prozentzahlen³⁾ mittels der Produkt-Moment-Korrelationstechnik bzw. dem t-Test zu verrechnen, muß abgelehnt werden. Differenzwerte sind meistens nicht normal-verteilt, eine entsprechende Überprüfung der Verteilungsform wurde von HUBER nicht vor-

1) Vgl. G. EIGLER: "Prüfung von Lehrprogrammen", in: ZeF, 2, 1968, 3, S. 126

2) Vgl. K.J. KLAUER, a.a.O., S. 3

3) Vgl. G. HUBER, a.a.O., S. 40, 47 und 48

genommen. Die Verwendung von Prozenten setzt weiter äquidistante Skalenpunkte und einen absoluten Nullpunkt voraus, bei dem von HUBER gewählten informellen Test zur Erfassung der abhängigen Variable kann davon ausgegangen werden, daß diese Bedingungen nicht - oder nur zum Teil - erfüllt sind. Vorbildlich scheint uns das Vorgehen von ISSING und SCHELLENBERG zu sein, die von vornherein auf parameterbezogene Prüfverfahren verzichteten. Dies sollte grundsätzlich gefordert werden, solange nicht der Nachweis der Normalverteilung der abhängigen Variable erbracht ist. Gerade bei informellen Tests kann davon nicht ohne weiteres ausgegangen werden.

4.00 Beschreibung und Analyse der Programmierten Unterweisung und des herkömmlichen Unterrichts

4.10 Lernpsychologische Grundlagen der Programmierten Unterweisung

Es ist zweifellos das Verdienst von SKINNER, auf seiner Lerntheorie aufbauend, eine Revolutionierung des konventionellen Unterrichts angestrebt zu haben, wobei er sowohl die Ablehnung der herkömmlichen Methoden des Unterrichts als auch seinen Alternativvorschlag lerntheoretisch fundierte. Im Gegensatz zu den vielen Versuchen der Pädagogik, den bestehenden Unterricht zu verändern bzw. durch einen anderen zu ersetzen, liegt hier zum ersten Mal ein systematischer Ansatz vor, welcher auf einer operational formulierten und experimentell (jedenfalls in Tierversuchen, aber auch darüberhinausgehend in humanpsychologischen Experimenten) abgesicherten Theorie aufbaut¹⁾. Kein Lerntheoretiker, nicht einmal ein Mann wie

1) SKINNERS Lerntheorie besteht allerdings in weitaus geringerem Ausmaß als dies bei anderen Lerntheoretikern, insbesondere HULL, der Fall ist, aus einem komplizierten System zwingend abgeleiteter und sorgfältig definierter theoretischer Aussagen. "Skinner's 'System' hat die Entwicklung der Forschung weniger beeinflusst als seine Experimente, in denen er virtuos wirksame Steuerungsfaktoren des Verhaltens zu demonstrieren versteht" (K. FOPPA: "Lernen, Gedächtnis, Verhalten", Köln-Berlin 1965, S. 358). Diese Gering-schätzung der theoretischen Formulierung zugunsten experimenteller Manipulation hat allerdings, wie wir später sehen werden, zur Folge, daß viele Begriffe, insbesondere der Begriff der Verstärkung, zwar formal eindeutig aber inhaltlich unklar beschrieben werden.

THORNDIKE¹⁾, hat für die Praxis des Unterrichtens durch seine Lerntheorie so entscheidende Impulse gesetzt wie dies bei SKINNER der Fall ist²⁾. Dies geschah einmal durch seine Theorie der operanten Konditionierung, welche in sehr grundsätzlicher Weise wesentliche Aspekte jeder unterrichtlichen Aktivität darstellt. Noch bedeutender wurde jedoch die praktische Umsetzung seiner Lerntheorie in die Konzeption des programmierten Textbuches³⁾ und in weiterem Sinne der Programmierten Unterweisung überhaupt⁴⁾. Ein Großteil der Kritik an der Programmier-

-
- 1) In diesem Zusammenhang ist die Veröffentlichung von E.L. THORNDIKE "Psychologie der Erziehung", Darmstadt 1970, zu erwähnen. THORNDIKE gibt, insbesondere im zweiten Teil seines Buches, eine Vielzahl von Hinweisen, wie Ergebnisse der psychologischen Forschung für die Erziehung fruchtbar gemacht werden können. Im Gegensatz zu SKINNERS Ansatz handelt es sich jedoch hier nicht um einen systematischen Versuch, von einer bestimmten Lerntheorie ausgehend, die Praxis des herkömmlichen Unterrichts zu kritisieren.
 - 2) Wie sehr die auf SKINNER zurückgehende Programmierte Instruktion gegenwärtig die Diskussion der Neugestaltung des Unterrichtswesens bestimmt, wird u.a. aus der Veröffentlichung von W. ZIFREUND (Hrsg.): "Schulmodelle, Programmierte Instruktion und Technische Medien", München 1968, ersichtlich.
 - 3) J.G. HOLLAND and B.F. SKINNER: "The Analysis of behavior", New York 1961
 - 4) So schreibt R. FREUDENSTEIN: "Der Programmierte Unterricht und seine lerntheoretischen Grundlagen", in W. KLAFKI u.a. (Hrsg.): "Erziehungswissenschaft 2", Frankfurt 1970: "...Der programmierte Unterricht gilt heute, 15 Jahre nach seiner Einführung, unangefochten als eine Antwort auf die Fragen, von denen Günter Klotz in dem zitierten Beleg behauptet, sie seien für das Leben des Menschen künftig existenznotwendig." (S. 229)

ten Unterweisung wie auch die Vielzahl der im Sinne der SKINNERschen Theorie unbrauchbaren Programme, die überzogen positiven und übertrieben negativen Erwartungen haben ihren Grund allzuoft in einer ungenügenden Kenntnis der amerikanischen, insbesondere aber der SKINNERschen Lerntheorie. Im folgenden werden wir daher diejenigen Aspekte der SKINNERschen Lerntheorie kurz berichten, welche für die Konzeption und Diskussion der programmierten Unterweisung von Bedeutung sind und gleichzeitig bei der vergleichenden Bewertung von herkömmlichem und programmiertem Unterricht berücksichtigt werden müssen. Insbesondere werden wir den Begriff der Verstärkung (reinforcement) und die Rolle der ausdrücklichen aktiven Reaktion (overt response) für den Lernprozeß diskutieren, da unsere Arbeit an dieser Stelle einen direkten experimentellen Beitrag leisten will.

4.11 Das Prinzip der Verstärkung

Im Gegensatz zu anderen lerntheoretischen Systemen¹⁾ steht bei SKINNER der Begriff der Verstärkung im Zentrum seiner Theorie²⁾. Ohne Verstärkung gibt es kein Lernen,

1) Hier ist insbesondere G. GUTHRIE zu nennen, der davon ausgeht, daß das einmalige gemeinsame Auftreten von Reiz und response eine endgültige und nicht verstärkbare Verbindung schafft.

2) An dieser Stelle muß ausdrücklich vermerkt werden, daß SKINNER die Verstärkung nicht als Erklärungsprinzip (wie z.B. HULL) versteht, sondern lediglich als einen beschreibbaren Aspekt einer experimentellen Anordnung, der die Reaktionswahrscheinlichkeit eines Verhaltens erhöht.

gelernte Verhaltensweisen sind Wirkreaktionen, denen ein verstärkender Reiz zeitlich unmittelbar gefolgt ist, welcher einen Zuwachs an Stärke der die Verstärkung auslösenden Reaktion bewirkt hat. "Wenn auf eine Wirkreaktion ein verstärkender Reiz folgt, so ergibt sich ein Zuwachs an Stärke"¹⁾. "Die Wirkreaktion gewinnt für das Verhalten eine Bedeutung und nimmt eine identifizierbare Form an, wenn sie auf die Umwelt in dem Sinne einwirkt, daß ein verstärkender Reiz hervorgerufen wird"²⁾. Aufgrund sorgfältiger Messungen, wobei als Maß für die Stärke einer Wirkreaktion ihr Widerstand gegen die Extinktion genommen wurde, kam SKINNER zu der Vermutung, daß bei Vorhandensein von nur einer Verstärkung ein unmittelbarer Übergang auf ein Maximum an Auftrittswahrscheinlichkeit einer Wirkreaktion erreicht wird³⁾. Wenn zur Erlernung einer Wirkreaktion in der Regel trotzdem eine Vielzahl von Verstärkungen notwendig ist, so deswegen, weil an einer äußerlich beobachtbaren Wirkreaktion in Wirklichkeit mehrere Wirkreaktionen beteiligt sind. SKINNER entwarf nun im Verlauf seiner experimentellen Studien sog. Verstärkungsprogramme (schedules of reinforcement), d.h. Systeme von planmäßig verabreichten Verstärkungen und untersuchte deren Auswirkung auf die Schnelligkeit der Konditionierung eines gewünschten Verhaltens sowie

1) B.F. SKINNER: "The behavior of organism: an experimental analysis", New York 1938, S. 21

2) B.F. SKINNER, a.a.O., S. 22

3) Vgl. B.F. SKINNER: "Science and human behavior", New York 1953, S. 68

dessen Resistenz gegen Extinktion. Besonders erwähnenswert sind dabei in unserem Zusammenhang die "fixed-interval-reinforcement" (Festintervallverstärkung: hier tritt nach festgelegten Intervallen regelmäßig eine Verstärkung ein, und die "variable-interval-reinforcement" (aperiodische Verstärkung: hier wird ein mittleres Intervall des Verstärkungsabstands zugrunde gelegt, eine Reaktion wird dabei z.B. durchschnittlich einmal in 5 Min. verstärkt, ohne daß festgelegt ist, wann diese Verstärkung innerhalb dieses mittleren Intervalls zu erfolgen hat). Vernaltensformen, welche mit einem aperiodischen Programm aufgebaut worden sind, haben sich gegen Extinktion als resistenzfähiger erwiesen¹⁾, wobei diese Befunde nur in Grenzen generalisierbar sind²⁾. Weiter können noch die "fixed-ratio-reinforcement" und die "variable-ratio-reinforcement" erwähnt werden. Hier gelten analog dieselben Unterschiede, nur daß in diesem Fall anstelle eines zeitlich definierten Intervalls, innerhalb dessen eine Verstärkung verabreicht wird, Reaktionsquoten des lernenden Organismus zugrunde gelegt werden. Die "fixed-ratio-reinforcement" und "fixed-inter-

1) Vgl. W. CORRELL: "Programmiertes Lernen und schöpferisches Denken", München 1965, S. 13

2) So scheint es uns außerordentlich problematisch zu sein, die einschlägigen Befunde für die programmierte Unterweisung fruchtbar zu machen. SKINNER selbst hat diesen Versuch auch nie unternommen. Einen indirekten Hinweis darauf ergibt die Wahl des Verstärkerprogramms in seinem programmierten Textbuch "The Analysis of Behavior".

vall-reinforcement" haben den Nachteil, daß nach der verstärkten Reaktion eine Art Reaktionspause eintritt, der Organismus hat hier gewissermaßen die feste Reaktionsquote (bzw. das zeitliche Intervall), auf welche hin jeweils eine Verstärkung erfolgt, gelernt und tendiert in den Zwischenpausen zur Inaktivität¹⁾. Die entsprechenden Programme können auch gemischt werden, wobei die Frage, welches Verstärkungsprogramm optimal ist, auch im Hinblick auf die Art des zu erlernenden Verhaltens gestellt werden muß²⁾. In jedem Fall muß bei der Handhabung solcher Verstärkerprogramme darauf geachtet werden, daß eine unmittelbare zeitliche Nachbarschaft zwischen dem vom Versuchsleiter (Unterrichtenden) gewünschten Verhalten und dem durch dieses ausgelösten verstärkenden Reiz besteht, da mit wachsendem zeitlichen Abstand die Wirkung des Verstärkers immer geringer wird³⁾.

Im Zusammenhang mit dem Prinzip der Verstärkung muß noch auf die Unterscheidung zwischen primärer und sekundärer Verstärkung hingewiesen werden. "Ein Reiz, der ursprünglich keine Verstärkung bewirkt, kann durch wiederholtes gemeinsames Auftreten mit einem Verstärker selbst zu einem solchen werden"⁴⁾, d.h. durch Konditionierung erwirbt ein Reiz die Fähigkeit, selbst eine solche herbeizuführen. Wichtig für die sekundären Verstärker ist,

1) Vgl. W. CORRELL, a.a.O., S. 12

2) Im Zusammenhang mit der Forschung über Verstärkerprogramme sei insbesondere auf die Veröffentlichung von C.B. FERSTER and B.F. SKINNER: "Schedules of reinforcement", New York 1957, hingewiesen.

3) Vgl. B.F. SKINNER: "Die Wissenschaft vom Lernen und die Kunst des Lehrens", in: W. CORRELL (Hrsg.): "Programmiertes Lernen und Lehrmaschinen", Braunschweig 1966, S. 73 f.

4) E.R. HILGARD: "Theorien des Lernens", Stuttgart 1970, S. 144

daß sie ausgelöscht werden, sofern sie wiederholt auf eine Reaktion hin erfolgen, für welche letztlich keine primäre Verstärkung erfolgt. Unbesehen dieser Tatsache ist die sekundäre Verstärkung jedoch unabhängig und unspezifisch, d.h. sie kann auch andere Verhaltensweisen verstärken als jene, über welche sie ursprünglich mit dem primären Verstärker verbunden war. Im Bereich des schulischen Lernens spielt vor allem die Verstärkung von sprachlichem Verhalten eine zentrale Rolle. Hier wird nach SKINNER primär auf sekundäre Verstärker zurückgegriffen: "Sprachlichem Verhalten, etwa Benennen von Objekten mit den ihnen zukommenden Namen, pflegt von den Zuhörern eine generalisierte Verstärkung zuteil zu werden, wenn sie auf indirekte Weise erkennen lassen, ob sie diese Aussage verstehen und ihr zustimmen"¹⁾. Schließlich muß noch erwähnt werden, daß Verstärkungen nicht nur dazu dienen, neues Verhalten aufzubauen sondern darüber hinaus notwendig sind, um zu gewährleisten, daß bereits gelerntes Verhalten beibehalten wird. Ungeachtet der Tatsache, daß die der SKINNERschen Theorie zugrunde gelegten Experimente vorzüglich mit Säugtieren und Vögeln durchgeführt worden sind, beansprucht SKINNER auch für das menschliche Lernen die Gültigkeit der Bedeutung der Verstärkung für das Erlernen von Verhaltensweisen, er mißt ihr letztthin genau denselben

1) E.R. HILGART, a.a.O., S. 145

Stellenwert zu¹⁾). Wenn wir uns an dieser Stelle zunächst einmal SKINNER vorbehaltlos anschließen, dann impliziert die Funktion des Begriffs der Verstärkung im Kontext des SKINNERschen Systems für den unterrichtenden Lehrer folgende Konsequenzen:

- a) jedes gewünschte Schülerverhalten muß, sobald es auftritt, sofort verstärkt werden;
- b) ein gewünschtes Schülerverhalten muß mehrfach und möglichst in jeweils anderem Kontext verstärkt werden um einen optimalen Lernerfolg zu gewährleisten;
- c) beim Aufbau komplizierter Verhaltensformen, zu deren Erlernung eine Vielzahl von Verstärkungen erforderlich ist, sollte jeweils das optimale Verstärkungsprogramm gewählt werden;
- d) es kann auch auf sog. Sekundär-Verstärker wie individuelles Lob, soziale Anerkennung usw. zurückgegriffen werden. Dabei sollten diese sekundären Verstärkersysteme regelmäßig mit primären assoziiert werden;
- e) bereits gelerntes Verhalten muß weiterhin - wenn auch seltener - verstärkt werden um es gegen Extinktion resistenzfähig zu machen.

1) So schreibt SKINNER in seinem Aufsatz "The science of learning and the art of teaching", in: W. CORRELL (Hrsg.): "Programmiertes Lernen und Lehrmaschinen", Braunschweig 1966, S. 71: "... Bei all diesen Arbeiten war die Gattung des Organismus überraschend unwesentlich. Zwar waren die untersuchten Lebewesen alle Wirbeltiere aber dennoch von sehr verschiedener Art. Vergleichbare Ergebnisse erzielten wir mit Tauben, Ratten, Hunden, Affen, Kindern und in letzter Zeit auch mit psychisch Erkrankten ... Trotz großer phylogenetischer Unterschiede zeigt der Lernprozeß dieser Lebewesen erstaunlich ähnliche Eigenschaften."

An dieser Stelle muß noch eine Bemerkung über die Wirkung von negativen Verstärkungen und damit über die Auswirkung von Bestrafung auf den Lernprozeß gemacht werden. Nach SKINNER hat eine Bestrafung lediglich zur Folge, daß die relative Reaktionshäufigkeit des negativ verstärkten Verhaltens niedergehalten wird, eine Schwächung der Wirkreaktionstendenz findet dagegen nicht statt. Sobald die Einwirkung der negativen Verstärkung eine Zeitlang ausgesetzt hat, erholt sich die unterdrückte Reaktion spontan. Reaktionen können also durch Bestrafung unterdrückt, nicht aber ausgelöscht werden¹⁾. Wenn ein Organismus unerwünschtes - und das heißt in der Schule fehlerhaftes - Verhalten zeigt, wird dies am besten dadurch beseitigt, daß es überhaupt nicht - weder positiv noch negativ - verstärkt wird, dagegen erwünschte Reaktionstendenzen umso konsequenter²⁾. Daraus folgt indirekt, daß bei jedem Konzept der Verhaltensformung darauf geachtet werden sollte, fehlerhafte Reaktionen grundsätzlich zu vermeiden. Überzeugende experimentelle Nachweise über die Wirkung negativer Verstärker finden sich bei ESTES³⁾. Hieraus muß folgende pädagogische Konsequenz gezogen werden:

- f) für den Prozeß der Verhaltensformung ist die negative Verstärkung ungeeignet, darüber hinaus hat die aversive Lehrerkontrolle sogar stark negative Auswirkungen⁴⁾.

1) Vgl. E.R. HILGARD, a.a.O., S. 160

2) Vgl. W. CORRELL, a.a.O., S. 14

3) Vgl. E.R. HILGARD, a.a.O., S. 160

4) Vgl. B.F. SKINNER: "Why Teachers Fail", in: W. CORRELL (Hrsg.): "Zur Theorie und Praxis des Programmierten Lernens", Darmstadt 1969, S. 93 f.

- g) fehlerhafte Reaktionen sollten soweit als möglich vermieden werden.

Zusammenfassend kann vermerkt werden, daß verstärkende Reize unabdingbare Voraussetzung für den Zuwachs an Stärke einer Wirkreaktion sind. Die Handhabung der Verstärkungsprogramme ist das entscheidende unterrichtliche Problem, welches sich für den Lehrer stellt. Genau an dieser Stelle setzt die massive Kritik SKINNERS am herkömmlichen Unterricht ein¹⁾, genau an dieser Stelle setzt aber auch die Kritik an SKINNER ein²⁾. Unsere Aufgabe wird es sein, die Praxis der Programmkonstruktion im Hinblick auf die Handhabung der Verstärker kritisch zu analysieren und vergleichend mit dem herkömmlichen Unterricht zu betrachten.

-
- 1) Vgl. B.F. SKINNER: "The science of learning and the art of teaching", in: W. CORRELL (Hrsg.): "Programmiertes Lernen und Lehrmaschinen", Braunschweig 1966, S. 74
- 2) So behauptet E.M. KRANICH in seiner Veröffentlichung "Pädagogische Projekte und ihre Folgen", Stuttgart 1969, S. 12, daß das Prinzip der Verstärkung zwar ein wesentlicher Lernfaktor sei, der "aber nur für eine niedrige, ganz von außen geleitete Form des Lernens" zutraf. A. WITTE kommt aufgrund von eigenen Untersuchungen zu folgendem Schluß: "Die sogenannten Lernverstärkungen des Programms haben weder direkt motivierende Wirkung, noch vermögen sie den Verschleiß einer anfänglich bestehenden Motivation zu verhindern" (A. WITTE: "Experimentelle Beiträge zur Entwicklung des kombinierten Unterrichts", in: "Lehrmaschinen in kybernetischer und pädagogischer Sicht", Stuttgart-München 1965, S. 191). Auch in russischen und amerikanischen Veröffentlichungen wird Kritik am Prinzip der Verstärkung im Zusammenhang mit der Programmierten Unterweisung laut. (Vgl. dazu A.N. LEONTJEW und P.J. GALPERIN: "Die Theorie des Kenntniserwerbs und der Programmierte Unterricht", in: W. CORRELL (Hrsg.): "Zur Theorie und Praxis des Programmierten Lernens", Darmstadt 1969, S. 110 und A.A. LUMSDAINE: "Probleme beim programmierten Unterricht", in: W. CORRELL (Hrsg.): "Zur Theorie und Praxis des Programmierten Lernens", Darmstadt 1969, S. 84).

4.12 Das Prinzip der ausdrücklichen aktiven Reaktion (overt response)

"Der Schüler muß seine Antwort selbst formulieren (im Sinne der Konstruktions-Antwort-Methode) statt sie aus einer Reihe von Auswahlmöglichkeiten zu wählen, wie in einem Auswahl-Antwort-Apparat. Ein Grund dafür ist, daß er aus dem Gedächtnis etwas reproduzieren und nicht nur etwas wiedererkennen soll. Er soll nämlich eine Antwort geben und gleich darauf erkennen, daß sie richtig war. Ein anderer Grund dafür ist, daß gutes Auswahl-Antwort-Material auch plausible falsche Antworten enthalten muß, die in dem schwierigen Prozeß der Formung des Verhaltens fehl am Platze sind, weil sie unerwünschte Verhaltensformen verstärken"¹⁾. Kennzeichnend für die extrem behavioristische Ausrichtung SKINNERs ist, daß er im Gegensatz zu HULL, TOLMAN u.a. die Verwendung von intervenierenden Variablen bzw. hypothetischen Konstrukten ablehnt. Die Beobachtung von Reiz und response ist für ihn die einzige sichere Quelle der Erkenntnis, die Erforschung der Bedingungen des tierischen und menschlichen Lernens kann also nur auf die unmittelbare Beobachtung gegründet werden. Für den "radikalen Operationalismus"²⁾ ist es kennzeichnend, daß er sogar die verhaltensauslösende Wirkung eines Reizes für einen Organismus ablehnt, sofern derartige Auslöser nicht unmittelbar feststellbar sind³⁾. Für SKINNER steht das operante Verhalten, die sog. Wirkreaktion, im Vordergrund menschl-

-
- 1) B.F. SKINNER: "Lehrmaschinen", in: W. CORRELL (Hrsg.): "Programmiertes Lernen und Lehrmaschinen", Braunschweig 1966, S. 41
 - 2) R.S. WOODWORTH: "Experimental psychology", New York 1954, S. 116
 - 3) Vgl. K. EYFERTH: "Lernen als Anpassung des Organismus durch bedingte Reaktionen", in: R. BERGIUS (Hrsg.): "Lernen und Denken", Hdb. d. Psych., Bd. 1, 2 Göttingen 1964, S. 99

chen Verhaltens. Gelernt wird, wenn eine Wirkreaktion einen verstärkenden Reiz zur Folge hat. Gedachte Antworten sind nicht direkt beobachtbar, sie haben daher im Kontext der Theorie der operanten Konditionierung keinen Platz, da diese Theorie ausschließlich an beobachtbarem spontanen Wirkverhalten gewonnen wurde¹⁾. Von hier aus wird es verständlich, daß das Wirkverhalten und damit die äußerlich unmittelbar beobachtbare "Reaktion" für SKINNER neben der Verstärkung den wesentlichen Teil jedes Lernprozesses darstellt²⁾. Ein solches Wirkverhalten muß geradezu herbeigeführt werden, damit es verstärkt und somit der Lernprozeß überhaupt vollzogen werden kann. In diesem Sinne ist N.A. CROWDER, ein Kritiker der SKINNERschen Konzeption, zu verstehen, wenn er schreibt: "Bei strenger Anwendung der linearen Theorie ist die Frage, wie der Schüler da-

-
- 1) Diese Beschränkung auf direkt beobachtbare Daten und damit auf das einfache S-R Schema wurde in jüngerer Zeit im Zusammenhang mit der Diskussion der programmierten Unterweisung häufig kritisiert. So schreiben A.N. LEONTJEW und P.J. GALPERIN, a.a.O., S. 111: "Wer den eigentlichen Aneignungsvorgang aus dem betrachteten Schema ausschließt, beschränkt sich darauf, Prozesse auszulösen oder aufgrund ihres Ergebnisses auszuwählen, anstatt ihren Ablauf zu steuern. Auch bei den kleinsten aufeinanderfolgenden Schritten wird nicht die Auseinandersetzung des Schülers mit dem Stoff, sondern lediglich ihr Ergebnis korrigiert oder bekräftigt."
 - 2) Vgl. N.A. CROWDER: "Unterschiede zwischen linearem und verzweigtem Programmieren", in: W. CORRELL (Hrsg.): "Zur Theorie und Praxis des programmierten Lernens", Darmstadt 1969, S. 152

zu veranlaßt wird, die richtige Antwort zu geben, irrelevant; wichtig ist, die Antwort überhaupt hervorzurufen, so daß sie belohnt und so 'gelernt' werden kann"^{1,2}).

Aus dem oben Ausgeführten muß, wiederum unter vorläufigem Verzicht auf eine kritische Stellungnahme, folgende unterrichtliche Konsequenz gezogen werden:

h) die Schüler sind ständig zu ausdrücklichen, aktiven Reaktionen anzuhalten.

4.13 Das Prinzip des "shaping of behavior"

Nach SKINNER läßt sich die Komplexität tierischen und menschlichen Verhaltens dadurch verstehen, daß man es als Ketten- oder Verlaufsmuster von Unterscheidungen von Reizen und Reaktionsdifferenzierungen auffaßt. Gerade diese von SKINNER gewählte Betrachtungsweise von tierischem und menschlichem Verhalten hat sich im Verlauf der praktischen Anwendung seines lerntheoretischen Systems - z.B. im Zusammenhang mit Tierdressuren - als ausgesprochen nützlich erwiesen. Hierauf baut auch die Konstruktion des SKINNERSchen Linearprogramms auf.

1) N.A. CROWDER, a.a.O., S. 152

2) In diesem Sinne ist auch die Kritik von A.N. LEONTJEW und P.J. GALPERIN, a.a.O., S. 111, zu verstehen: "Ein programmierter Unterricht, der den Aneignungsprozeß nicht berücksichtigt und nur nach dem Schema Stimulus-Reaktion-Verstärkung abläuft, läßt das entscheidende Glied - den Prozeß des Kenntniserwerbs - unberücksichtigt."

"Der ganze Prozeß des Sicherwerdens auf jedem Gebiet muß in eine sehr große Zahl sehr kleiner Schritte aufgeteilt werden, und mit jeder Durchführung eines Schrittes muß eine Verstärkung verbunden werden ..."¹⁾ ...

"Indem man jeden einzelnen folgenden Schritt so klein wie möglich hält, kann die Häufigkeit der Verstärkung ein Maximum erreichen, während die möglichen negativen Folgen des Fehlermachens auf ein Minimum reduziert werden ..."²⁾. Dieses Prinzip des shaping soll im folgenden anhand der Erlernung einer komplexen motorischen Verhaltensweise kurz aufgezeigt werden: Zunächst wird ein relativ undifferenziertes Wirkverhalten (z.B. Heben des Armes) verstärkt. Sobald diese Verhaltensform gelernt ist, wird nur noch ein differenzierteres Wirkverhalten (z.B. Heben des rechten Armes) verstärkt. Auf diese Weise wird sukzessiv eine Kette von immer differenzierterem Wirkverhalten stufenweise verstärkt, bis schließlich das gewünschte Endverhalten ausgebildet ist (z.B. Heben des rechten Armes bei gleichzeitigem Spreizen der Finger der rechten Hand). Dasselbe Prinzip kann für die Unterscheidung von Signalen angewandt werden³⁾.

-
- 1) B.F. SKINNER: "Die Wissenschaft vom Lernen und die Kunst des Lehrens", in: W. CORRELL (Hrsg.): "Programmiertes Lernen und Lehrmaschinen", Braunschweig 1966, S. 78
 - 2) B.F. SKINNER, a.a.O., S. 79
 - 3) Auf diesem Hintergrund ist die Herstellung der Kurve der Dunkeladaptation einer Taube zu sehen, wie sie von F. RATLIFF and D.S. BLOUGH: "Behavioral studies of visual processes in the pigeon", Report of Contract N5ori-07663, Psychological Laboratories. Harvard Univ., Sept. 1954, berichtet wird.

Lernen kann also grundsätzlich als eine Verhaltensformung durch Verwendung sukzessiver Näherungsschritte bezeichnet werden. Das Fehlen solcher durchdachter Programme der Verhaltensformung durch sukzessive, möglichst kleine Näherungsschritte ist ein weiterer Punkt, an dem SKINNERS Kritik am herkömmlichen Unterricht ansetzt. Nach SKINNER kann auch das komplexeste Verhaltensmuster (z.B. Verständnis für historische Zusammenhänge) nach dem Prinzip des shaping gelernt werden. Gerade an dieser Stelle setzt von verschiedenen Seiten massive Kritik ein¹⁾, wir werden diese im weiteren Verlauf unserer Arbeit zu besprechen haben.

Aus dem Prinzip des shaping muß folgende unterrichtliche Konsequenz gezogen werden:

- 1) der unterrichtete Stoff muß in kleinen, sukzessiv aufeinander aufbauenden Lernschritten dargeboten werden. Die Größe der Lernschritte muß so bemessen sein, daß falsche Reaktionen möglichst auf ein Minimum reduziert werden.

4.14 Die Rolle des "einsichtigen Lernens" im System von SKINNER

In den Schlagwortverzeichnissen von SKINNERS Büchern tritt, obwohl SKINNER Lerntheoretiker ist, der Begriff "Einsicht" nicht auf. Nach KELLER und SCHOENFELD ist Einsicht im SKINNERSchen System mit schnellem Lernen gleichzusetzen, welches "auf der Grundlage a) der Ähn-

1) Vgl. H. SKOWRONEK: "Psychologische Grundlagen einer Didaktik der Denkerziehung", Hannover 1968, S. 151

lichkeit des vorliegenden mit einem früher einmal gelösten Problem oder b) der Einfachheit des Problems" erklärt werden kann¹⁾. Unbekannte Problemsituationen werden durch schrittweise Verstärkung immer differenzierterer Wirkreaktionen gelernt, nicht aber etwa durch plötzliches Umstrukturieren im Sinne der gestaltpsychologischen Konzeption. Häufig wird von den Befürwortern des "einsichtigen Lernens" die Darbietung schwieriger Lernschritte befürwortet und dies von dem "mechanischen Lernen einfacher Fakten" streng abgegrenzt. Als Argument wird u.a. ins Feld geführt, daß die Schüler nur in der Konfrontation mit schwierigen Lernschritten das "Denken" lernen würden und daß zudem das, was leicht gelernt werde, auch leicht vergessen werde. In diesem Zusammenhang schreibt SKINNER: "Es gibt keinen überzeugenden Grund dafür, daß das, was leicht gelernt wird, rascher vergessen werden soll. ... Die übliche Befürwortung des 'schwierigen' Stoffes ist der Hinweis, daß wir mehr als Fakten lehren wollen. Der Schüler soll herausgefordert werden, er soll 'denken' lernen. Dieses Argument ist oft wenig mehr als der Ausdruck einer verworrenen Vorstellung, aber ohne Zweifel ist es wahr, daß Vorlesungen und Schulbücher oft unangemessen und absichtlich in die Irre führen. Aber wozu? Welche Art von 'Denken' soll der Schüler lernen, indem er sich durch schwieriges Material hindurchkämpft? Es ist wahr, daß diejenigen, die unter schwierigeren Bedingungen lernen, bessere Schüler sind, aber sind sie besser deswegen, weil sie Schwierigkeiten überwunden haben, oder überwinden sie diese Schwierigkeiten, weil sie besser sind? Unter dem Vorwand, Denken zu lehren, führen wir schwierige und verworrene Situationen herbei und berufen uns dabei auf

1) E.R. HILGARD, a.a.O., S. 165

die Schüler, die damit erfolgreich fertig werden. ... Ein vernünftigeres Verfahren wäre es, das Verhalten, das wir 'Denken' nennen zu analysieren und es dann entsprechend zu schulen. Für ein derartiges Programm läge bereits Material auf den Gebieten der Logik, Mathematik, Methodologie und Psychologie vor."¹⁾ Ungeachtet der Frage, ob es möglich ist, mit solchen Argumenten die Problematik des einsichtigen Lernens abzutun²⁾, wollen wir an dieser Stelle zunächst die unterrichtliche Konsequenz daraus ziehen:

- j) auch "Denken" kann durch entsprechende Programme der Verhaltensformung in kleinen Lernschritten gelernt werden.

4.20 Begriffsklärung und Analyse des linearen Programmtypus

Wir werden im folgenden den von SKINNER entwickelten linearen Programmtypus behandeln und analysieren, inwieweit und auf welche Weise die von uns aus seiner

1) B.F. SKINNER: "Lehrmaschinen", in: W. CORRELL (Hrsg.): "Programmiertes Lernen und Lehrmaschinen", Braunschweig 1966, S. 58

2) Ganz besondere Bedeutung gewinnt in diesem Zusammenhang die Arbeit von H. AEBLI: "Psychologische Didaktik", Stuttgart 1968, der unter Rückgriff auf die genetisch orientierte Intelligenztheorie von J. PIAGET interessante Beiträge zur Gestaltung des Unterrichtsstoffes und dessen Darbietung liefert. Einen erschöpfenden Überblick über psychologische Beiträge, welche z.T. in deutlichem Widerspruch zu SKINNERs Konzeption stehen, bietet H. SKOWRONEK in seiner Veröffentlichung "Psychologische Grundlagen einer Didaktik der Denkerziehung", Hannover 1968. Zweifellos muß im Hinblick auf diese und andere Veröffentlichungen die SKINNERsche Vorstellung über eine Optimalisierung der schulischen Lernsituationen kritisch überdacht werden.

Lerntheorie abgeleiteten unterrichtlichen Konsequenzen realisiert worden sind. Nicht lineare, von der SKINNERschen Konzeption abweichende Programmtypen werden nicht behandelt, einmal, da wir unserer eigenen Untersuchung ein Linearprogramm zugrunde gelegt haben und zum anderen, weil es nicht in der Absicht dieser Arbeit liegen kann, diesbezüglich eine erschöpfende Übersicht zu geben¹⁾.

Schematisch kann das Konstruktionsprinzip des auf SKINNER zurückgehenden linearen Programmtypus wie folgt umschrieben werden: Dem Schüler wird zunächst ein kurzer und vollständiger Informationsteil geboten. Nachdem er diesen aufgenommen hat, wird ihm die Aufgabe gestellt, eine sich unten anschließende unvollständige Information zu ergänzen, indem er den fehlenden Teil schriftlich formuliert und an der entsprechenden Stelle einsetzt bzw. er wird aufgefordert, eine Frage schriftlich zu beantworten, deren Lösung i.allg. aus einem Reaktionswort besteht. Die vom Schüler verlangte Antwort bezieht sich dabei unmittelbar auf den vorgegebenen vollständigen Informationsteil. Nachdem der Schüler die Antwort schriftlich festgehalten hat, kann er (i.allg. auf der folgenden Seite) die von ihm gegebene Antwort mit der richtigen vergleichen und dabei Übereinstimmung bzw. Nichtübereinstimmung registrieren. An dieser Stelle endet ein Lernschritt, ihm folgt ein nächster Lernschritt, der sich bezüglich seines vollständigen Informationsteils an den vorangegangenen Lernschritt unmittel-

1) In diesem Zusammenhang sei auf die Arbeit von J.P. LYSAGHT und C.M. WILLIAMS: "Einführung in die Unterrichtsprogrammierung", München-Wien 1967, verwiesen.

bar anschließt. Sämtliche das Programm bearbeitende Schüler haben - unbesehen ihrer Begabung und ihres individuellen Lerntempos - sämtliche vorgegebenen Lernschritte in einer vorher festgelegten, für jeden Schüler gleichen Reihenfolge zu durchlaufen¹⁾.

- a) Die von SKINNER für so wichtig gehaltene sofortige Verstärkung wird dadurch gewährleistet, daß der Schüler unmittelbar nach jeder aktiv formulierten Antwort die Übereinstimmung seiner Reaktion mit der vom Lerninhalt sachlich verlangten Antwort feststellen kann. Zwischen der vom Schüler gegebenen Antwort und der Bestätigung liegt keine weitere stoffbezogene Schüleraktivität, so daß gewährleistet ist, daß tatsächlich die gewünschte Verhaltensweise verstärkt wird.
- b) Die Forderung nach mehrfacher Verstärkung einer angestrebten Verhaltensweise wird dadurch erfüllt, daß diese in verschiedenen, i.allg. unmittelbar aufeinander folgenden Lernschritten eingeübt wird, wobei der Kontext, in welchem die angestrebte Verhaltensweise eingeübt wird, laufend variiert wird²⁾. An dieser Stelle muß vermerkt werden, daß diese Einübung der gewünschten Reaktion in verschiedenem Kontext häufig auch in SKINNER-Programmen nicht mehr ausschließlich auf dem Boden der S-R Verstärkungstheorie sondern nach der S-R Kontinguitätstheorie von GUTHRIE³⁾ erfolgt, d.h. das gewünschte Reaktions-

1) Auf diesen Tatbestand bezieht sich die Bezeichnung "extern gesteuertes Programm".

2) Vgl. J.P. LISAUGHT und G.M. WILLIAMS, a.a.O., S. 119

3) E.R. GUTHRIE: "The psychology of learning", New York 1952

wort wird nur ein- oder zweimal verstärkt, erscheint dann aber in darauffolgenden Lernschritten in einem jeweils variierten assoziativen Umfeld¹⁾.

- c) Die Forderung nach einem optimalen Verstärkungsprogramm wird nach unserem Wissen von keinem Programmator, auch nicht von SKINNER, erfüllt. Den SKINNER-Programmen ist grundsätzlich die "fixed-ratio-reinforcement" zugrunde gelegt, d.h. also ein starres Verstärkungsprogramm. Dies steht in deutlichem Gegensatz zu den Aussagen innerhalb der Theorie von SKINNER. Wenngleich dieser davon ausgeht, daß das jeweils optimale Verstärkungsprogramm von Stoff zu Stoff neu überprüft und festgelegt werden sollte, hat es sich doch als deutliche Tendenz herausgebildet, daß die variablen Verstärkungsprogramme sich als grundsätzlich überlegen erweisen^{2,3)}.
- d) Der regelmäßige Rückgriff auf primäre Verstärkersysteme innerhalb eines Programms ist aus technischen Gründen nicht möglich, wenngleich auch bei Program-

-
- 1) Vgl. A.A. LUMSDAINE: "Probleme beim Programmierten Unterricht", in: W. CORRELL (Hrsg.): "Zur Theorie und Praxis des Programmierten Lernens", Darmstadt 1969, S. 84, und J.G. HOLLAND and B.F. SKINNER: "Analysis of Behavior", New York 1961, S. 41-45
 - 2) Vgl. W. CORRELL, a.a.O., S. 13
 - 3) Angesichts der Tatsache, daß die Komplexität des vom Schüler bei der Durcharbeitung des Programms verlangten Verhaltens ungleich viel größer ist als jene motorischen Reaktionen von Tieren, an denen die Effektivität der verschiedenen Verstärkerprogramme überprüft worden ist, ergeben sich - wie wir bereits oben erwähnt haben - grundsätzliche Schwierigkeiten der Übertragung gerade dieser Befunde auf die Programmkonstruktion.

men der Inhalt des verstärkenden Reizes variiert wird. Im allgemeinen dient als verstärkender Reiz die Darbietung der richtigen Lösung, manchmal werden aber auch noch Zusatzreize gegeben, welche primären Verstärkersystemen wenigstens "näher" stehen¹⁾. So werden z.B. manchmal zusätzlich zur Darbietung der richtigen Lösung aufmunternde Worte, wie "das hast Du gut gemacht" oder "arbeite weiter so aufmerksam", gegeben. Streng genommen handelt es sich - was die Entstehungsbedingungen anbelangt - ebenfalls um sekundäre Verstärkersysteme, da primär verstärkende Reize i.allg. auf die Befriedigung von Hunger, Durst und Sexualität bzw. die Vermeidung von Schmerz beschränkt sind. Da die Auswirkung der sekundären Verstärkersysteme, also das Lernen des Stoffes, häufig zur Folge hat, daß primäre Verstärkersysteme indirekt wirksam werden (so führt z.B. ein guter Lernerfolg zu einer guten Schulnote, welche ihrerseits vom Vater des Schülers zum Anlaß genommen werden kann, diesem eine finanzielle Zuwendung zu machen, welche in unmittelbarem assoziativen Bezug zur Befriedigung primärer Motive steht), kann davon ausgegangen werden, daß der zur Erhaltung der sekundären Verstärker notwendige Rückgriff auf primäre Verstärker in ausreichendem Maße gewährleistet ist. Hier scheint uns übrigens eine wichtige Lehreraufgabe vorzuliegen. Er hat die Möglichkeit, das im allgemeinen sachlich ausgerichtete Verstärkersystem der Unterrichtsprogramme in Richtung auf verstärkende Reize wie "soziale Aner-

1) Die Wirksamkeit der Programmverstärker wird von mancher Seite aus als ausgesprochen gering angesehen. Folgerichtig verweisen diese Autoren auch auf die "besseren" Möglichkeiten des lehrergesteuerten Unterrichts. (Vgl. in diesem Zusammenhang A. WITTE, a.a.O., S. 191).

kennung" u.a. auszuweiten¹⁾.

- e) Dem Übungsaspekt wird in vielen Programmen durch eigens eingeschobene Übungseinheiten, d.h. eine Sequenz von Lernschritten, welche ausschließlich der Verfestigung des Gelernten dienen, Rechnung getragen. Es wäre auch daran zu denken, eigens Übungsprogramme zu entwerfen, welche z.B. den Stoff eines Jahreskurses für die einzelnen Fächer den Lernschritten zugrunde legen würden, wobei als Lernziel ausschließlich die Verfestigung von bereits Gelerntem gewählt würde. Im übrigen gilt das bereits unter b) Ausgeführte.
- f) Negative Verstärkungen sind - im Gegensatz zum verzweigten Programmtypus von CROWDER - in SKINNER-Programmen grundsätzlich nicht zu finden. Die üblichen aversiven Lehrerkontrollen²⁾ entfallen bei Lehrprogrammen grundsätzlich.

-
- 1) SKINNER wird häufig vorgeworfen, er würde den Begriff der Verstärkung in unscharfer Weise gebrauchen. (Vgl. N. CHOMSKY: "Review of B.F. Skinner, Verbal Behavior". Language 35, 1959, S. 26-58). H. KAY: "General introduction to teaching machine procedures", in: K. AUSTWICK (Ed.): "Teaching machines and programming", Oxford 1964, differenzierte zwischen der rein sachlichen (confirmation) und der emotional getönten Verstärkung (reinforcement). Im deutschsprachigen Bereich untersuchte H. BRAUNE die verschiedene Auswirkung dieser beiden Verstärkerarten auf die Lernleistung (H. BRAUNE: "Untersuchung über die Wirkungsweise verschiedener Bestätigungen (reinforcement; confirmation) in einem verzweigten Programm unter Berücksichtigung differentiell-psychologischer Lernvariablen", in: ZeF, 3, 1969, 3 und 4). Es scheint sicher zu sein, daß bei Zugrundelegung eines optimalen Verstärkerprogramms beide Aspekte berücksichtigt werden müssen, wobei beim Programmierten Lernen dem Lehrer in erster Linie die Handhabung der emotional getönten Verstärker zukommen dürfte.
 - 2) Vgl. B.F. SKINNER: "Warum Lehrer versagen", in: W. CORRELL (Hrsg.): "Zur Theorie und Praxis des Programmierten Lernens", Darmstadt 1969, S. 93 ff.

- g) Die Forderung nach der Vermeidung fehlerhafter Reaktionen wird dadurch erfüllt, daß der Schwierigkeitsgrad der verlangten Antworten so gehalten wird, daß möglichst jeder Schüler auf Anhieb zur richtigen Reaktion gelangt. Für die Konstruktion von SKINNER-Programmen wird dabei häufig die auf SKINNER zurückgehende Forderung gestellt, daß ein Lernschritt von 95 % der dem Programm zugrunde gelegten Adressatengruppe auf Anhieb richtig gelöst werden sollte¹⁾. Um fehlerhafte Reaktionen zu vermeiden, werden weiter die verschiedensten Lernhilfen zum Einsatz gebracht. Diese reichen von einer geringen Informationskomponente der einzelnen Lernschritte²⁾ bis zur Technik der Bereitstellung adäquater Antwortmöglichkeiten³⁾.
- h) Programme vom "Typ-SKINNER" vermeiden grundsätzlich die Antwortauswahlmethode, wie sie in den verzweigten Programmen verwendet wird. Die Schüler werden in jedem Lernschritt grundsätzlich zur aktiven Formulierung des Reaktionswortes bzw. der Reaktionswortsequenz gezwungen. Die gebräuchlichste Technik besteht darin, die Schüler vorgegebene Lücken ausfüllen zu lassen, wobei darauf geachtet werden muß, daß das verlangte

-
- 1) An dieser Stelle muß vermerkt werden, daß diese Bedingung streng genommen von den im Handel befindlichen Programmen in den allerseltensten Fällen erfüllt wird, im deutschsprachigen Raum ist uns überhaupt kein solches Programm bekannt.
- 2) Vgl. J.P. LYSAUGHT und C.M. WILLIAMS, a.a.O., S. 117 und S. 119
- 3) Vgl. J.P. LYSAUGHT und C.M. WILLIAMS, a.a.O., S. 120 ff. und D.J. KLAUS: "Die Programmierung des Selbstunterrichts", in: W. CORRELL (Hrsg.): "Programmiertes Lernen und Lehrmaschinen", Braunschweig 1966, S. 257 ff.

Reaktionswort in einem direkten Bezug zum jeweiligen Lehrziel steht. Gerade diese Bedingung ist bei schlechten Programmen nicht erfüllt, hier müssen zwar auch Lücken ausgefüllt werden, das verlangte Reaktionswort ist jedoch für die angestrebte Verhaltensendform ohne Belang¹⁾.

- i) Die Forderung nach Unterteilung in kleine, sukzessiv einander zugeordnete Lernschritte wird bei SKINNER-Programmen formal durchweg erfüllt²⁾. Um sicher zu sein, daß auch wirklich kleine Lernschritte aufeinanderfolgen, werden die unter g) geschilderten Konstruktionsprinzipien der Lernschrittkonstruktion zugrunde gelegt. Fraglich ist allerdings, ob die kleinen Lernschritte immer so angeordnet sind, daß sie vom übergeordneten Lernziel her inhaltlich optimal gereiht sind. Gerade dieser sehr wichtige Aspekt wird bei der Programmkonstruktion häufig vernachlässigt. Zu lösen ist er nur über den Weg sorgfältiger didaktischer Studien unter Zuhilfenahme des Konzepts der strengen operationalen Lernzielformulierung, wie es insbesondere von R.F. MAGER beschrieben und begründet wurde³⁾. Unbesehen des Bemühens um operationale Lernzielformulierung wurde zwischenzeitlich eine Vielzahl von Programmierungstechniken entwickelt, welche sich mit der Frage der optimalen Abfolge der einzelnen Lernschritte

1) Vgl. D.J. KLAUS, a.a.O., S. 263

2) Vgl. J.P. LYSAUGHT und C.M. WILLIAMS, a.a.O., S. 119 und S. 123

3) Vgl. R.F. MAGER: "Preparing Objectives for Programmed Instruction", San Francisco, 1962

befassen¹⁾.

- j) Die Frage, ob durch kleine Lernschritte "einsichtiges Lernen" gewährleistet werden kann, bzw. ob der Schüler dadurch "denken" lernt, ist heftig umstritten²⁾. Es wird der jeweiligen Interpretation der einzelnen Lehrprogramme überlassen bleiben müssen, ob diese Bedingung als erfüllt bzw. nicht erfüllt angesehen werden kann. Soweit diese Thematik im Zusammenhang mit der Programmkonstruktion diskutiert wird, finden sich die entsprechenden Beiträge in der unter i) aufgeführten Literatur³⁾. Wir wollen uns an dieser Stelle SKINNER anschließen, der mit Recht darauf hinweist, daß die Behauptung, man könne bei kleinen Lernschritten nicht "denken" lernen, bislang nicht überzeugend verifiziert worden ist.

-
- 1) In diesem Zusammenhang sind insbesondere folgende Autoren zu nennen:
J. EVANS, R. GLASER and L. HOMME: "The Ruleg System for the Construction of Programmed Verbal Learning Sequences", Department of Psychology, University of Pittsburgh 1960
T.F. GILBERT: "Mathetics: The Technology of Education, I. The General System", TOR Education, Inc., New York 1961
F. MECHNER: "Programming for automated instruction", New York: Basic Systems, Inc., 1961
R. GLASER: "Some research problems in automated instruction: instructional programming and subject-matter structure", in: E. COULSON (Ed.): "Programmed Learning", New York 1962
- 2) Optimistisch äußert sich in diesem Zusammenhang W. CORRELL, a.a.O., S. 68: "Es besteht aber kein Zweifel daran, daß grundsätzlich auch schöpferisches Denken programmierbar ist und daß entsprechende Untersuchungen vom Standpunkt der Verhaltenspsychologie im Laufe der Zeit vorliegen werden". Äußerst skeptisch äußert sich dagegen u.a. J.S. BRUNER (vgl. dazu J.S. BRUNER: "The art of discovery". Harv. Ed. Rev., 1961, 31, S. 26 ff).
- 3) Vgl. auch D.J. KLAUS, a.a.O., S. 266. Hier versucht der Autor anhand von Lernschritten aufzuzeigen, daß diese so konstruiert werden können, daß die Schüler selbsttätig etwas entdecken und d.h. wohl selbständig "denken".

Eine interessante Variation der üblichen SKINNER-Programme ist das Programmierungsmodell der konversationalen Verkettung von J.A. BARLOW¹⁾. Diese Variation unterscheidet sich von dem oben beschriebenen Programmtypus insbesondere dadurch, daß die Antwortverkettung innerhalb des Programms intensiver gestaltet ist. Dies wird dadurch erreicht, daß die richtige Antwort auf den vorhergehenden Reiz in den nächstfolgenden Informationsteil des Lernschrittes eingebaut ist. Das gewünschte Antwortverhalten wird also nicht isoliert dargeboten um erst dann zum nächsten Lernschritt vorzugehen, vielmehr ist das gewünschte Antwortverhalten selbst Bestandteil des nächsten Lernschrittes²⁾. Ob der Schüler das gewünschte Antwortverhalten gezeigt hat, kann er daraus ersehen, ob er seine für den vergangenen Lernschritt gegebene Antwort in dem neuen Lernschritt vorfindet. Das gewünschte Reaktionswort kann dabei - i.allg. wird so verfahren - z.B. gesperrt gedruckt werden, so daß der Schüler keine Schwierigkeit hat, die Richtigkeit seiner vorangegangenen Antwort zuverlässig festzustellen.

Schließlich ist noch zu erwähnen, daß die strenge externe Steuerung, d.h. also die festgelegte Zahl und Richtung der von den Schülern - und zwar von allen gleichermaßen - zu durchlaufenden Lernschritte bei den sog. modifizierten linearen Programmen, den linearen Programmen mit Nebenlinien und den linearen Programmen mit Kriteriumseinheiten etwas aufgelockert ist. Dies

-
- 1) Vgl. J.A. BARLOW: "Conversational Chaining in Teaching Machine Programs", Earlham College, Richmond 1960
 - 2) Hinter diesem Programmierungsmodell steht zweifellos GUTHRIE'S Kontinguitätstheorie. Dies wird u.a. auch daraus ersichtlich, daß keine ausdrückliche Reaktion verlangt wird.

geschieht dadurch, daß dem Schüler ermöglicht wird, bei einem bestimmten Antwortverhalten einige Lernschritte zu überspringen bzw. daß von ihm verlangt wird, bei Nichtvorhandensein des gewünschten Verhaltens bestimmte Lernschritte zu wiederholen. Weiter bietet sich im Falle von Nebenlinien die Möglichkeit, gute Schüler mit Zusatzinformation zu versorgen. Im Falle des Vorliegens von Kriteriumseinheiten wird festgestellt, ob der Schüler über ein bestimmtes Wissen verfügt, welches Voraussetzung für das Überspringen einiger Lerneinheiten im Hauptprogramm darstellt¹⁾. Alle die hier geschilderten Programme bzw. Programmierungsmodelle sind im wesentlichen nur Variationen des ursprünglich von SKINNER entwickelten linearen Programms, allen gemeinsam muß das Prinzip der externen Steuerung, des aktiven Antwortverhaltens²⁾, der Verstärkung sowie das Prinzip der Verhaltensformung durch kleinste Lerneinheiten (shaping) zugrunde gelegt werden.

4.30 Allgemeine Erörterung des herkömmlichen Unterrichts

Im Gegensatz zur Programmierten Unterweisung, welche unmittelbar aus einer psychologischen Lerntheorie entwickelt wurde und deren Funktionsprinzipien, jedenfalls bezüglich der formalen Aspekte, ziemlich eindeutig definiert sind, umfaßt der sog. herkömmliche Unterricht, welcher den Vergleichsstudien als unabhängige Variable zugrunde gelegt wird, eine Vielzahl von Unterrichtsformen, von denen keine unmittelbar lernpsychologisch be-

1) Hier liegt bereits eine enge Nachbarschaft zu CROWDERS verzweigtem Programmtypus vor.

2) Das gilt nicht für das Modell von BARLOW.

gründet ist. Diese Unterrichtsformen reichen von den sog. darbietenden über die erarbeitenden zu den mittelbaren Unterrichtsformen¹⁾. Unter darbietender Unterrichtsform wird dabei eine Stoffvermittlung verstanden, welche vorzüglich auf dem Lehrervortrag bzw. der Lehrerdemonstration basiert. Die erarbeitenden Unterrichtsformen zeichnen sich dadurch aus, daß der Unterricht in Form eines vom Lehrer im Hinblick auf die Lehrziele gesteuerten Lehrer-Schüler-Gesprächs abläuft, die mittelbaren Unterrichtsformen umfassen Einzel- und Gruppenarbeit, wobei die Schüler sich mit den vom Lehrer zur Verfügung gestellten Arbeitsmitteln auseinandersetzen²⁾. Sämtliche hier geschilderten Unterrichtsformen entstanden aus der Wechselwirkung von pädagogischer Theorie und praktischer Schularbeit.

Im historischen Kontext werden im deutschsprachigen Raum die Unterrichtslehren von HERBART, GAUDIG und KERSCHENSTEINER besonders häufig diskutiert. Jeder dieser Autoren legte ein umfangreiches theoretisches Konzept vor, in welchem die Funktionsprinzipien des Unterrichts sowie deren pädagogisch-philosophische Begründung abgehandelt sind. So sehen HERBART und seine Anhänger das Ziel des Unterrichts in der Begriffsbildung, der Lehrer hat dabei die Aufgabe, "den Schüler durch ständige direkte Einwirkung, und zwar meist in Form der Frage, schrittweise von Stufe zu Stufe zu führen"³⁾, wobei die schematische Gliederung der Unterrichtsstunde "Vorbe-

1) Vgl. K. STÖCKER: "Neuzeitliche Unterrichtsgestaltung", München 1970, S. 8 f.

2) An dieser Stelle wird von der herkömmlichen Didaktik auch die programmierte Unterweisung eingeordnet. Vgl. K. STÖCKER, a.a.O., S. 282

3) Vgl. H. NOHL, E. WENIGER, G. GEISLER (Hrsg.): "Das Problem der Unterrichtsmethode", Kleine Pädagogische Texte, Heft 19, Weinheim, o.J., S. 4

reitung", "Darbietung", "Verknüpfung", "Zusammenfassung" und "Anwendung" umfaßt¹⁾. H. GAUDIG knüpfte an HERBART unmittelbar an und vertrat die Ansicht, "daß die herbartianische Unterrichtsmethode die Spontaneität der Schüler nicht zur Entfaltung kommen lasse"²⁾. Für GAUDIG war demgegenüber das Prinzip der "Freitätigkeit" des Schülers grundlegend für seine Unterrichtslehre. Dabei muß der Lehrer versuchen, die Schüler in die jeweiligen stoffbezogenen Arbeitstechniken einzuweisen. "Dem Lehrer muß die Methode, seine Zöglinge zur Methode zu führen, eigen sein"³⁾. Der formale Ablauf einer Unterrichtsstunde bei GAUDIG ist so geartet, daß der Schüler sich zunächst an einer Arbeit versucht bzw. der Lehrer eine Arbeit vorführt. Danach wird vom Lehrer die Technik erläutert, die Schüler versuchen nun diese unter zurückhaltender Anleitung des Lehrers nachzuvollziehen, die Technik wird solange eingeübt, bis sie beherrscht wird. Dem schließt sich das Stadium des freien Arbeitens an. Die für GAUDIG zentrale Phase der Selbsttätigkeit ist ebenfalls gegliedert. Der Schüler setzt sich dabei selbst das Ziel seiner Arbeit, er versucht das Arbeitsziel Schritt für Schritt zu realisieren, fällt an kritischen Punkten selbsttätig Entscheidungen, nimmt notwendige Korrekturen vor und beurteilt schließlich rückschauend das Ergebnis seiner Arbeit⁴⁾. KERSCHENSTEINER steht GAUDIG ziemlich nahe, hier werden die Akzente dahingehend verschoben, daß die Arbeit nicht als Methode sondern als Inhalt des Unterrichts betrachtet wird⁵⁾.

1) Vgl. H. NOHL, E. WENIGER, G. GEISSLER (Hrsg.); a.a.O., S.4

2) Vgl. H. NOHL, E. WENIGER, G. GEISSLER (Hrsg.); a.a.O., S.5

3) Vgl. H. NOHL, E. WENIGER, G. GEISSLER (Hrsg.); a.a.O., S.5

4) Vgl. H. NOHL, E. WENIGER, G. GEISSLER (Hrsg.); a.a.O., S.47 ff.

5) Vgl. H. NOHL, E. WENIGER, G. GEISSLER (Hrsg.); a.a.O., S.55 ff.

Wir wollen uns an dieser Stelle eine detailliertere Ausführung der KERSCHENSTEINERSchen Gedankengänge ersparen. Dieser kurze historische Ausschnitt¹⁾ soll nur aufzeigen, daß auch der herkömmliche Unterricht bezüglich seiner wesentlichen Funktionsprinzipien und seiner Zielsetzung definierbar ist, wenngleich weder die Ableitung dieser Unterrichtslehren aus der zeitgenössischen Philosophie bzw. Psychologie (wie bei HERBART) noch die Beschreibung der Funktionsprinzipien so eindeutig und theoretisch stringent sind, wie dies bei der Programmierten Unterweisung der Fall ist, jedenfalls solange diese sich auf SKINNER beruft²⁾. Dieser vorläufige Hinweis auf die Definierbarkeit des herkömmlichen Unterrichts scheint uns notwendig zu sein, da Vergleichsuntersuchungen, wie wir oben gesehen haben, häufig mit dem Argument abgelehnt werden, es sei nicht möglich, den

-
- 1) Der von uns gewählte historische Ausschnitt kann auch nicht annähernd einen Anspruch auf Vollständigkeit oder Repräsentativität erheben, dazu wäre es zumindest noch erforderlich, die in jüngerer Zeit besonders häufig diskutierten Konzepte von E. WENIGER und W. KLAFKI heranzuziehen (vgl. H. BLANKERTZ: "Theorien und Modelle der Didaktik", München 1969). In dem oben ausgeführten Problemzusammenhang scheint uns jedoch eine ausführliche Diskussion überflüssig zu sein.
 - 2) Vereinzelt wird im Zusammenhang mit der pädagogischen Diskussion der Programmierten Unterweisung versucht, gemeinsame Funktionsprinzipien zwischen den historisch überlieferten Unterrichtsformen und der Programmierten Unterweisung aufzuzeigen. Dieser Ansatz dürfte allerdings nicht viel mehr als einige Analogieschlüsse erwarten lassen, da sich der Vergleich auf formale Stilelemente beschränkt und die Rückbesinnung auf die lerntheoretischen Hintergründe der Programmierten Unterweisung nur ungenügend erfolgt. Vgl. E.S. GEISSLER: "Herbarts Theorie des 'kleinsten Schrittes' und der Programmierte Unterricht", Westermanns Pädagogische Beiträge, 18, 1966, S. 189-197, und L. MÜLLER: "Gaudigs freie geistige Schularbeit und das Programmierte Lernen", Schweizerische Lehrerzeitung, Zürich 111, 1966, S. 1484-1487.

herkömmlichen Unterricht ausreichend zu definieren¹⁾. Dabei werden im Zusammenhang mit dieser Ablehnung die z.T. schon weit entwickelten Methoden zur empirischen Erfassung des herkömmlichen Unterrichts nicht erwähnt²⁾. Sicher erfüllen Beschreibungen, wie sie im Rahmen der Unterrichtslehre üblich sind, längst nicht die strengen Anforderungen, welche an die Definition von unabhängigen Variablen im Rahmen eines Experiments gestellt werden müssen³⁾. Einmal abgesehen von dieser Tatsache wollen wir jedoch an dieser Stelle zunächst festhalten, daß es durchaus möglich ist, verschiedene herkömmliche Unterrichtsformen deutlich voneinander abzuheben. Im Anschluß an das unter 2.00 und 3.00 Vermerkte kann daher festgehalten werden, daß die häufig praktizierte Gepflogenheit, bei Vergleichsuntersuchungen einfach vom "herkömmlichen Unterricht" zu reden ohne ihn näher zu beschreiben, keinesfalls damit gerechtfertigt werden kann, daß er so und so nicht beschreibbar sei und daß man eben von einem "durchschnittlichen Unterricht" ausgehen müsse.

-
- 1) Vgl. K.J. KLAUER: "Über die Erfassung der Validität von Lehrprogrammen", in: pl 1, 1968, S. 2
 - 2) Vgl. D.M. MEDLEY and H.E. MITZEL: "Measuring Classroom Behavior by Systematic Observation", in: N.L. GAGE (Ed.): "Handbook of Research on Teaching", Chicago 1965 und F. WINNEFELD: "Pädagogischer Kontakt und Pädagogisches Feld", München-Basel 1965
 - 3) Wir sind uns darüber im klaren, daß die im Rahmen der pädagogischen Theorie zur Abgrenzung der verschiedenen Unterrichtslehren verwendeten Termini bei weitem nicht den Grad an Eindeutigkeit haben, welcher für die experimentelle Setzung von Variablen gefordert werden muß. Allerdings vertreten wir die Ansicht, daß aufgrund einer solchen Abgrenzung mittels einer weiterführenden detaillierten Analyse nach bestimmten Kriterien eine Unterrichtsbeschreibung erfolgen kann, welche den notwendigen Grad an Eindeutigkeit - wenigstens annäherungsweise - erreicht.

An dieser Stelle ergibt sich nun ein weiteres Problem. Es ist, wie wir oben gesehen haben, zwar möglich, den typischen Unterricht im Sinne HERBARTs vom typischen Unterricht im Sinne GAUDIGs abzuheben und dies gilt analog für weitere Unterrichtslehren und Unterrichtsformen, welche sich im Verlauf der Geschichte der Pädagogik herausgebildet haben. Nun muß aber davon ausgegangen werden, daß ein Lehrer in den seltensten Fällen einen herkömmlichen Unterricht praktiziert, der eindeutig einem Typ zugeordnet werden kann, es kann wohl grundsätzlich angenommen werden, daß sich in irgendeinem konkreten Unterricht typische Unterrichtsformen überlagern. Da zudem ein bestimmter herkömmlicher Unterricht nicht durch einen, sondern letztthin durch viele Lehrer repräsentiert wird, müssen wir darüber hinaus noch mit -z.T. starken - individuellen Abweichungen rechnen. Das hat zur Folge, daß zusätzlich differenziertere Kategorien entwickelt werden müssen, welche unbesehen der individuellen Abweichungen des Lehrerverhaltens für die praktizierte Unterrichtsform kennzeichnend sind. Wir werden daher im folgenden versuchen, typische Merkmale des unserer Untersuchung zugrunde gelegten herkömmlichen Unterrichts (bzw. der zugrunde gelegten herkömmlichen Unterrichte, da ja mehrere Lehrer an der Untersuchung beteiligt waren) herauszuarbeiten. Diese Merkmale müssen so formuliert sein, daß sie einmal für alle in unserer Untersuchung praktizierten Unterrichte gleichermaßen Gültigkeit haben und zum andern die Voraussetzung dafür schaffen, daß eine klare Unterscheidung zu anderen Formen des herkömmlichen Unterrichts hin möglich ist. Dies ist auch Voraussetzung dafür, daß wir später den Bereich der Generalisierbarkeit unserer Untersuchungsergebnisse sicherer abschätzen können. Schließlich sollen die herauszuarbeitenden typischen Merkmale des unserer Untersuchung zugrunde gelegten herkömmlichen Unterrichts als Grundlage

für den Vergleich mit den Funktionsprinzipien der programmierten Unterweisung - und hier speziell des linearen, extern gesteuerten Programms vom "Typ SKINNER" - dienen. Dieser ist Voraussetzung für die Formulierung und Absicherung unserer Hypothesen, daß dazu die bisherigen einschlägigen Untersuchungsergebnisse zusätzlich mit herangezogen werden müssen ist selbstverständlich.

4.31 Das frontalunterrichtliche Verfahren

Wir haben bei der Wahl unseres herkömmlichen Vergleichsunterrichts darauf geachtet, daß dieser soweit als möglich im Sinne einer experimentell gesetzten Variable definierbar ist. Dabei haben wir drei Wege zur Eingrenzung und Definition eingeschlagen. Zuerst haben wir aus der Vielzahl der möglichen Unterrichtsformen eine ausgewählt und dabei ausdrücklich darauf geachtet, daß andere Unterrichtsformen nicht mit einbezogen werden. Diesen Schritt wollen wir als Definition der Makrostruktur unseres Unterrichts bezeichnen. Zum anderen haben wir alle an der Untersuchung beteiligten Lehrer mit detaillierten Anweisungen versehen, an welche sie sich bei der praktischen Durchführung ihres Unterrichts zu halten hatten. Diesen Schritt wollen wir als die Sicherung der Mikrostruktur unseres herkömmlichen Unterrichts bezeichnen. Schließlich haben wir mittels Beobachtungsprotokollen, unter Zuhilfenahme von Tonbandaufnahmen, die Aktions- und Interaktionsformen von Lehrern und Schülern festgehalten. In diesem Kapitel wollen wir uns ausschließlich mit der Makrostruktur unseres herkömmlichen Unterrichts befassen, die detaillierten Arbeitsanweisungen für die Lehrer und die Protokollierung der Aktions- und Interaktionsformen

werden im Rahmen der Beschreibung unserer experimentellen Bedingungen abgehandelt. Schließlich muß noch gesagt werden, daß die Analyse unseres herkömmlichen Unterrichts unter dem Gesichtspunkt der dominanten Schüler- und Lehreraktivitäten sowie der Kontrolle des Ablaufs dieser Aktivitäten erfolgt.

Der Wahl unserer experimentell gesetzten Unterrichtsform ging eine Analyse der von den an unserer Untersuchung beteiligten Lehrern am häufigsten benutzten Unterrichtsformen voraus, wobei wir uns hier von Anfang an auf den Fachbereich Algebra beschränkten. Dabei ergab sich, daß alle Lehrer schwerpunktmäßig die sog. erarbeitende Unterrichtsform¹⁾ unter direkter Führung des Lehrers bevorzugten, wobei allerdings hinsichtlich des Einbezugs der darbietenden Unterrichtsform, d.h. des Vorhandenseins von Unterrichtspassagen, in denen der Lehrer vorzüglich Probleme dozierend vorträgt, beträchtliche Abweichungen vorlagen. Verschieden war auch das Ausmaß, in dem die Lehrer Allein- bzw. Partnerarbeit - insbesondere in den Übungsteilen des Unterrichtsverlaufs - mit einbezogen. Angesichts des deutlichen Überwiegens der erarbeitenden Unterrichtsform entschlossen wir uns, diese unserem Versuchsunterricht zugrunde zu legen. Die Lehrer wurden vor Beginn der experimentellen Phase ausdrücklich aufgefordert, auf längere dozierende Unterrichtspassagen sowie den Rückgriff auf mittelbare Unterrichtsformen wie Partner- und Gruppenarbeit und Stillarbeit zu verzichten. Das heißt, daß sowohl für die Problemhinführung, Problemstellung und Problemlösung als auch für die daran anschließende "Verfestigung durch Übung"

1) Vgl. K. STÖCKER: "Neuzeitliche Unterrichtsgestaltung", München 1970, S. 215 ff.

grundsätzlich das Lehrergespräch verwendet wurde¹⁾, wobei darauf zu achten war, daß möglichst alle Schülerbeiträge über den Lehrer liefen. Dies bedeutet, daß unterrichtsbezogene Gespräche zwischen Schülern ohne jeweiligen Lehrereingriff nach jedem einzelnen Schülerbeitrag vermieden werden sollten. Diese Einschränkung belastete die übliche unterrichtliche Situation nicht, da wir während unserer Vorversuche in keinem Fall solche Gesprächsformen feststellen konnten²⁾.

Wir können an dieser Stelle bereits folgende typischen Merkmale unseres herkömmlichen Unterrichts festhalten:

- a) Die unterrichtliche Aktivität wurde durch die Wahl des erarbeitenden Lehrer-Schüler-Gesprächs ziemlich gleichmäßig zwischen Lehrer und Schülern verteilt.
- b) Sämtliche Unterrichtsbeiträge seitens der Schüler liefen unmittelbar über den Lehrer, das unterrichtliche Geschehen unterlag also einer sehr starken Lenkung seitens des Lehrers.

Aus der starken Lenkung des Gesprächs durch den Lehrer geht hervor, daß das vom Lehrer jeweils erstrebte Unterrichtsziel von vornherein festliegt, Methode und Wegfindung sind keine gemeinsame Angelegenheit von Lehrern und Schülern, wie das beim sog. Unterrichtsgespräch im

1) Vgl. K. STÖCKER, a.a.O., S. 217 ff.

2) Dies scheint eine Eigentümlichkeit des Mathematikunterrichts zu sein, da in anderen Fächern, insbesondere in Geschichte und Deutsch, der Lehrer häufig über längere Zeit hinweg "außerhalb" der zwischen den Schülern ablaufenden Gespräche bleibt, nachdem er mit einem entsprechenden Impuls die Diskussion angeregt hat.

Gegensatz zum Lehrgespräch der Fall ist¹⁾. Um das Gespräch einzuleiten und ingang zu halten dienen dem Lehrer dabei die unterrichtlichen Mittel der Frage und des Impulses²⁾. Dabei dienen die Fragen insbesondere dazu, den Schüler dazu zu bringen, "einen vorliegenden Gegenstand unter einem bestimmten Gesichtspunkt zu betrachten"³⁾, der Impuls hat die Aufgabe, den Schüler zu stoffbezogener Aktivität aufzufordern und diese in eine vom Lehrer gewünschte Richtung zu lenken. Voraussetzung für die Fragestellung und Impulsgebung ist allerdings, daß beiden Informationen vorausgehen müssen, an welchen die Schüler anknüpfen können. Daraus ergeben sich weitere typische Merkmale unseres Unterrichts:

- c) Die Lehreraktivitäten bestehen vorwiegend in der Informationsgebung, dem Stellen von Fragen, der Impulsgebung und der Reaktion auf Schüleraktivitäten.
- d) Die Information dient lediglich dazu, Fragen bzw. Impulse vorzubereiten. Ein Informationsüberhang wird dadurch vermieden.

Die strenge Unterrichtslenkung durch den Lehrer hat weiter

-
- 1) Beim Lehrgespräch ist der Lehrer eindeutig führend, "die unterrichtliche Lage 'nicht offen', denn der Lehrer ist allein im Besitz der Wahrheit und führt die Schüler nach vorgefaßter Zielsetzung und Planung" (vgl. K. STÖCKER, a.a.O., S. 215). Diese Unterrichtsform kommt unserem Untersuchungsanliegen auch dadurch besonders entgegen, daß es sich um die gegenwärtig wohl "meistgebrauchte Form der Unterrichtsgestaltung" (vgl. K. STÖCKER, a.a.O., S. 215) handelt. Dies begünstigt die Generalisierbarkeit unserer Untersuchungsergebnisse.
 - 2) Entsprechend häufig finden sich die Bezeichnungen "Frageunterricht" und "Impulsunterricht" (vgl. K. STÖCKER, a.a.O., S. 217 f. und 223 f.).
 - 3) H. AEBLI: "Grundformen des Lehrens", München 1968, S. 88

zur Folge, daß er auf jede stoffbezogene Schüleraktivität, soweit es sich um ausdrückliche Gesprächsteiligungen handelt, reagieren muß.

- e) Der Lehrer reagiert auf jede an ihn gerichtete stoffbezogene Schüleraktivität.
- f) Die Schüler sind hinsichtlich einer Reaktion auf ihre stoffbezogene Aktivität auf den Lehrer angewiesen.

Da die Schüleraktivität durch Fragen und Impulse aufrechterhalten wird, werden die Schüler vorzüglich reaktiv agieren, d.h. Antworten geben.

- g) Die Schüleraktivität besteht vorzüglich im Geben von Antworten, sie ist also überwiegend reaktiv orientiert.

Die auf die Schüleraktivitäten bezogene Lehrerreaktion kann grundsätzlich in den Formen der Bestätigung, der Ablehnung oder der zusätzlichen Erteilung von Informationen erfolgen. Bestätigung und Ablehnung werden dabei überwiegen, sie gehen meistens der Erteilung einer neuen und weiterführenden Information voraus. Die Lehrerreaktionen auf Schüleraktivitäten erfolgen für alle Schüler vernehmbar, d.h. daß jeder Schüler, der sich aktiv am Unterricht beteiligt, der auf seine Tätigkeit hin erfolgenden Lehrerkontrolle und der Klassenresonanz ausgesetzt ist.

- h) Die Schülertätigkeit unterliegt ständig der Lehrerkontrolle. Die vom Lehrer gestellten Fragen können einmal kollektiv, d.h. an die gesamte Klasse erfolgen, zum anderen kann jeder einzelne Schüler jederzeit aufgerufen werden.
- i) Jeder Schüler muß ständig damit rechnen, zu einer Antwort aufgefordert zu werden.

Die Analyse des Schwierigkeitsgrades der einzelnen Fragen und damit in weiterem Sinne der Fragetechnik überhaupt gehört zwar nicht mehr in den oben benannten Beschreibungsbereich, da jedoch hieraus wesentliche unterrichtliche Konsequenzen resultieren, welche die Fragestellung unserer Arbeit unmittelbar betreffen, müssen wir die Fragetechnik zur Beschreibung unseres Unterrichts mit heranziehen. Eine nachträgliche, mit Hilfe von Tonbandaufnahmen durchgeführte Analyse derselben ergab, daß sie von der echt didaktischen Frage¹⁾ bis zu den von der modernen Unterrichtslehre i.allg. abgelehnten Ergänzungsfrage, Suggestivfrage und Entscheidungsfrage reichte²⁾. Dabei war deutlich die Tendenz zu erkennen, die Schüler mittels Fragen zu "selbständigem Denken" anzuregen. Bereits SKINNER hat darauf hingewiesen, daß sich der herkömmliche Unterricht dazu i.allg. schwieriger Fragen bedient³⁾. Die weit verbreitete Ansicht, der Frage die primäre Aufgabe eines "Denkanstoßes" zuzuschreiben und das damit häufig zusammenhängende Vorurteil, daß nur dort gedacht werden könne, wo die Frage entsprechend schwierig formuliert sei (und daher mit hoher Wahrscheinlichkeit mit falschen oder jedenfalls unvollständigen Antworten gerechnet werden muß), spiegelte sich auch in der Fragetechnik der im Rahmen un-

1) Vgl. H. AEBLI, a.a.O., S. 90

2) Vgl. K. STÜCKER, a.a.O., S. 220

3) Vgl. B.F. SKINNER: "Lehrmaschinen", in: W. CORRELL (Hrsg.): "Programmiertes Lernen und Lehrmaschinen", Braunschweig 1966, S. 58

serer Untersuchung unterrichtenden Lehrer wider.

- j) Die Lehrerfragen sind häufig so formuliert, daß sie falsche bzw. unvollständige Antworten zur Folge haben.

Hinweise, welche der Frage die primäre Aufgabe des "Denkanstoßes" zuweisen, finden sich gehäuft in den Lehrbüchern der Unterrichtslehre. In diesem Zusammenhang schreibt STÖCKER: "... meist wird der Lehrer durch seine weitgefaßten Fragen genauso wie durch seine Impulse erst die Denktätigkeit und Beobachtungsfähigkeit anregen"¹⁾. Es braucht an dieser Stelle nicht weiter ausgeführt zu werden, daß "weit" formulierte Fragen die Antwortrichtung nicht festlegen und daß damit die Wahrscheinlichkeit einer Fehlantwort beträchtlich steigt. Psychologisch fundiert ist die Warnung von AEBLI vor der zu eng gefaßten Lehrerfrage. Er meint, daß durch sie dem Schüler die Denkopoperationen und Erfassungsschemata vorgegeben werden und ihm somit spontane, eigenständige Denkarbeit abgenommen wird. Gleichzeitig fügt er aber hinzu: "... Indem man das Kind zu früh in die Selbständigkeit hinausstöße, bleibt es immer unselbständig, unsicher ..." ²⁾.

Wir glauben, mit dem oben Ausgeführten eine unterrichtliche Makrostruktur beschrieben zu haben, welche den in unserer Untersuchung praktizierten Unterricht von anderen Unterrichtsformen deutlich abhebt und gleichzeitig entscheidende Hinweise auf die Form der Lehrer- und Schüleraktivität sowie deren Kontrolle gibt. An dieser Stelle sei ausdrücklich vermerkt, daß wir uns

1) K. STÖCKER, a.a.O., S. 221

2) H. AEBLI, a.a.O., S. 95

darüber im klaren sind, daß die Frage der Lehreffektivität und die sich weiter im Zusammenhang mit unseren Hypothesen ergebenden Probleme in entscheidendem Maße von weiteren, hier nicht näher beschriebenen Kriterien abhängen. Es sei nur auf die zentrale Bedeutung der inhaltlichen Stoffaufbereitung, der F o r m u l i e - r u n g von Fragen und Problemen sowie des verwendeten Erziehungsstils hingewiesen. Wenngleich wir also einräumen, unseren herkömmlichen Vergleichsunterricht mit den oben beschriebenen Merkmalen nicht vollständig definiert zu haben, gehen wir doch davon aus, daß wir ihn - im Gegensatz zu anderen Veröffentlichungen - bezüglich wesentlicher Strukturmerkmale eindeutig beschrieben haben. Da deren Auswirkungen auf Lernprozesse aufweisbar und begründbar sind, haben wir nach unserer Auffassung eine bessere Grundlage zur Hypothesenbildung und Ergebnisinterpretation geschaffen als dies bei der üblichen Form der Vergleichsstudien, welche den herkömmlichen Unterricht kaum beschreiben, der Fall ist. Das gilt umso mehr, als uns weiter die detaillierten Vorschriften für das Lehrerverhalten sowie die empirische Kontrolle des Aktions- und Interaktionsgefüges unseres Versuchsunterrichts zur Verfügung stehen.

5.00 Vergleichende Betrachtung der Programmierten Unterweisung und des herkömmlichen Unterrichts

Wir werden uns im folgenden zunächst noch einmal kurz vergegenwärtigen, welche Vorteile die Programmierte Unterweisung gegenüber dem herkömmlichen Unterricht für sich in Anspruch nimmt. Danach werden wir die Frage überprüfen, inwieweit dieser Anspruch zu Recht besteht, wobei wir zu diesem kritischen Vergleich insbesondere unseren experimentell gesetzten herkömmlichen Vergleichsunterricht heranziehen werden. Ziel dieses Kapitels ist es, die Hypothesen unseres experimentellen Teils vorzubereiten.

An dieser Stelle sei noch vermerkt, daß wir dem Themenkreis "Verstärkung" besonders viel Raum widmen werden, da gerade von hier aus seitens SKINNER das Konzept der Programmierten Unterweisung theoretisch begründet wird.

5.10 Das Prinzip der Verstärkung

- a) Es ist unzweifelhaft ein Vorteil der Programmierten Unterweisung, daß jeder einzelne Schüler nach vollzogenem Wirkverhalten sofort verstärkt wird (vgl. dazu 4.11 a) und 4.20 a)), sofern wir davon ausgehen, daß eine sofortige Verstärkung für den Lernprozeß optimal ist¹⁾. Angesichts der diesbezüglichen Übereinstimmung der einschlägigen Arbeiten, welche im Rahmen der behavioristischen Lernforschung veröffentlicht worden sind, haben wir keinen Grund daran zu

-
- 1) D. AUBLE und E. MECH ("Response Strength in a Classroom Task Related to Forward Delay in Reinforcement", in: Journ. educ. psych., 45, 1954, S. 175 ff.) berichten zwar, daß in ihrem Experiment die sofortige Verstärkung sich der verzögerten nicht als überlegen erwiesen habe, dieses Experiment ist jedoch methodisch angreifbar (vgl. hierzu: W. CORRELL: "Pädagogische Verhaltenspsychologie", München-Basel 1967, S. 79) und kann im Kontext einer theoretischen Diskussion erst dann diskutiert werden, wenn wenigstens eine Wiederholungsstudie vorliegt. Im tierpsychologischen Bereich liegen mehrere Untersuchungen vor, welche sich mit der Frage der Bedeutung der sofortigen Verstärkung befassen, zusammenfassend kann zu diesen Untersuchungen gesagt werden, daß eine sofortige Verstärkung zwar nicht grundsätzlich für den Aufbau von Verhalten notwendig ist, daß verzögerte Verstärkungen jedoch nur dann wirksam werden, wenn der Organismus in der Verzögerungszeit seine Aufgabenhaltung beibehält (vgl. W. CORRELL, a.a.O., S. 78). Da auf der anderen Seite jedoch eine Vielzahl von Arbeiten vorliegen, welche überzeugend die Vorteile der sofortigen Verstärkung aufzeigen, scheint uns die Schlußfolgerung, "daß nach allen bisher vorliegenden Untersuchungen zu diesem Problem die sofortige Verstärkung gegenüber der verzögerten Verstärkung den Vorteil der Eindeutigkeit hat" (W. CORRELL, a.a.O., S. 79) richtig zu sein.

zweifeln¹⁾. Entsprechende Ergebnisse liegen auch seitens PAWLOW und seiner Schüler für das Signallernen vor²⁾. Diese unmittelbare und jeden Schüler betreffende Verstärkung ist bei der von uns gewählten Form des herkömmlichen Vergleichsunterrichts keinesfalls gewährleistet, darüber hinaus ist in der Tat keine Form des herkömmlichen, also lehrerbezogenen Unterrichts in der Lage, diese seitens der psychologischen Lerntheorien erhobene Forderung zu erfüllen. Wenngleich in unserem Vergleichsunterricht - im Gegensatz zu anderen herkömmlichen Unterrichtsformen - die Lehrer- und Schüleraktivität relativ gleichmäßig verteilt ist (vgl. 4.31 a)), alle ausdrücklich geäußerten Schüleraktivitäten über den Lehrer laufen (vgl. 4.31 b)), der Lehrer auf jede an ihn gerichtete Schüleraktivität reagieren muß (vgl. 4.31 e)) und die ausdrücklich geäußerte Schüleraktivität ständig der Lehrerkontrolle unterliegt (vgl. 4.31 h)), ist die individuelle und sofortige Verstärkung jeweils nur auf denjenigen Schüler beschränkt, der zu einem bestimmten Zeitpunkt im Verlauf der unterrichtlichen Situation mit dem Lehrer Kontakt aufnimmt. Zwar werden sich auf eine Lehrerfrage bzw. einen Lehrerimpuls hin i.allg. mehrere Schüler melden, aber selbst im günstigen Fall werden nur zwei oder drei aufgerufen und gelangen dadurch zur Möglichkeit einer aktiven Reaktion. Sofern der Lehrer mehreren Schülern Gelegenheit zur Antwort auf ein und dieselbe Frage gibt, wird er alle diejenigen,

1) Vgl. K. FOPPA: "Lernen, Gedächtnis, Verhalten", Köln-Berlin 1965, S. 98 ff.

2) Vgl. K. FOPPA, a.a.O., S. 28 ff.

welche vor dem zuletzt aufgerufenen Schüler geantwortet haben, sofern ihre Antwort richtig war, nicht sofort verstärken, sondern er wird noch die Antwort des zuletzt aufgerufenen Schülers abwarten. Dies liegt daran, daß der Lehrer im Falle der Antwortbestätigung des zuerst aufgerufenen Schülers das Aufrufen weiterer Schüler unterlassen wird, da er davon ausgeht, daß die einmal vollzogene Bekanntgabe der richtigen Lösung (Verstärkung) für die übrigen Schüler die Voraussetzung dafür schafft, ihre eigene, nicht ausdrücklich geäußerte Lösung auf Übereinstimmung bzw. Nichtübereinstimmung mit der richtigen Lösung zu überprüfen. Lernpsychologisch gesehen wird durch das Verzögern der Bekanntgabe der richtigen Lösung zwischen einem Schülerverhalten und dessen Verstärkung eine relativ große, die Wirkung des verstärkenden Reizes reduzierende bzw. auslöschende Zeitspanne geschaffen. Verstärkt der Lehrer, um diesen Nachteil zu vermeiden, dagegen einen Schüler sofort nach dessen Reaktion, werden alle übrigen Schüler (von SKINNER her argumentiert) nicht verstärkt, da sie ja nicht zu einer aktiven Reaktion gelangen. In jedem Fall aber wird die weitaus überwiegende Mehrzahl der Schüler auf eine Lehrerfrage hin überhaupt nicht verstärkt. Dies aber hat zur Konsequenz, daß pro Unterrichtseinheit auf den einzelnen Schüler relativ wenig Verstärkungen entfallen¹⁾. An dieser Stelle muß nun eine deutliche Einschränkung gemacht werden. Es ist für die amerikanische - und hier insbesondere

1) Vgl. B.F. SKINNER: "Die Wissenschaft vom Lernen und die Kunst des Lehrens", in: W. CORRELL (Hrsg.): "Programmiertes Lernen und Lehrmaschinen", Braunschweig 1966, S. 74

für die SKINNERsche - Lernpsychologie kennzeichnend, daß die Frage der Wirksamkeit von Verstärkungen wohl an einzelnen Organismen untersucht worden ist, nicht jedoch unter den Bedingungen der Gruppensituation, wie sie in einer Klasse besteht. Gerade SKINNER hatte eine Vorliebe für sog. Fallstudien. So schreibt er in seinem Artikel "A case history in scientific method": "Ein anderes Mißgeschick befreite mich von der mechanisierten Statistik und brachte mich zu einer noch intensiveren Konzentration auf den Einzelfall".¹⁾ Ohne die SKINNERschen Erfolge bezüglich individueller Tierdressuren in Frage stellen zu wollen, müssen wir an dieser Stelle vermerken, daß die Möglichkeit einer "kollektiven Verstärkung" nach unserem Wissen von SKINNER im Kontext seiner Lerntheorie nicht diskutiert wurde. Dies ist zwar von seinem methodischen Ansatz her verständlich, beantwortet jedoch nicht die Frage, ob "kollektive Verstärkung" möglich ist und welchen Bedingungen sie im einzelnen unterliegt. Wir vertreten hier die Ansicht, daß SKINNERs Begriff der Verstärkung von seinem methodischen Vorgehen her determiniert und daher bezüglich seiner Aussagekraft entsprechend eingeschränkt ist. Für ihn liegt Verstärkung nur dann vor, wenn einem Wirkverhalten in möglichst unmittelbarer zeitlicher Folge ein verstärkender Reiz folgt bzw. dieser vom Wirkverhalten hervorgerufen wird. Dabei versteht SKINNER unter Wirkverhalten ein ausdrückliches, für jedermann sichtbares Verhalten eines Organismus. Den Bereich der nicht sichtbaren inneren Aktivität schließt SKINNER von seinen Forschungsbemühungen aus²⁾. Demgegenüber vertreten wir die Meinung, daß

- 1) Vgl. B.F. SKINNER: "Fallstudie zur wissenschaftlichen Methode in der Psychologie", in: W. CORRELL (Hrsg.): "Programmiertes Lernen und Lehrmaschinen", Braunschweig 1966, S. 130
- 2) SKINNER meint zwar, daß "kein Tatbestand und kein Vorgang, der irgendwie zur Klärung der Fragen beitragen kann", abgelehnt werden darf, "bloß weil er subjektiv oder geistig ist" ("Fünfzig Jahre Behaviorismus", in: W. CORRELL: "Programmiertes Lernen und Lehrmaschinen", Braunschweig 1966, S. 110). Er fährt jedoch fort: "Die Gegebenheiten jedoch, die im bedeutend erscheinen lassen, müssen wirksam erforscht und formuliert werden" (B.F. SKINNER, a.a.O., S. 110). D.h. aber, daß Wege gefunden werden müssen, welche es ermöglichen, diese "inneren Prozesse" sichtbar zu machen. Nur so ist "Inneres" dem wissenschaftlichen Zugriff erschließbar (vgl. B.F. SKINNER, a.a.O., S. 100). In seiner Veröffentlichung "Verhaltenspsychologische Analyse des Denkprozesses" (in: B.F. SKINNER und W. CORRELL: "Denken und Lernen", Braunschweig 1967) räumt er zwar ein, daß jedes Verhalten "sich auf eine 'verinnerlichte' oder 'verdeckte' Ebene zurückziehen" kann, "wenn nur die gleichen Verstärkungszusammenhänge aufrechterhalten bleiben; dies ist der Fall, wenn die Verstärkung entweder automatisch erfolgt oder durch das nachfolgende offene Verhalten bewirkt wird" (S. 26 f.), auf der anderen Seite steht jedoch die eindeutige Definition des operativen Verhaltens, welches dadurch bestimmt ist, daß es auf die Umwelt einwirkt, und d.h. wohl in irgendeiner Form beobachtbar sein muß. Selbst wenn wir davon ausgehen, daß - angesichts der oben erwähnten automatischen Verstärkung - von SKINNER ausgesprochen so etwas wie "kollektive Verstärkung" zugestanden wird, würde SKINNER die Wirksamkeit derselben für schulische Lernprozesse als sehr gering veranschlagen, da die Schule sich damit befaßt "dem Kind eine große Zahl von Reaktionen von einer bestimmten Art zu vermitteln", diese Reaktionen jedoch "alle sprachlicher Art" sind (B.F. SKINNER: "Die Wissenschaft vom Lernen und die Kunst des Lehrens", in: W. CORRELL (Hrsg.): "Programmiertes Lernen und Lehrmaschinen", Braunschweig 1966, S. 72) und "verdecktes Verhalten" bestenfalls als s u b v o k a l e s verbales Verhalten bezeichnet werden könnte.

es zum Wirksamwerden eines verstärkenden Reizes nicht notwendig ist, daß diesem ein äußerlich sichtbares Wirkverhalten vorausgeht, vielmehr muß davon ausgegangen werden, daß "nur gedachtes", also "innerlich erfolgtes Schülerverhalten" ebenfalls verstärkt werden kann.

Neben der Einschränkung des Prinzips der individuellen Verstärkung aufgrund der Möglichkeit einer "kollektiven Verstärkung", d.h. der Erhöhung der Auftrittswahrscheinlichkeit von "verdecktem" Verhalten muß auch die Frage erörtert werden, inwieweit trotz zeitlicher Verzögerung von verstärkenden Reizen diese im Bereich humanpsychologischer Lernprozesse wirksam werden können. Die von uns oben zugestandene Tatsache, daß die Wirkung sofortiger Verstärkung im Gegensatz zur zeitlich verzögerten eindeutiger erwiesen ist, schließt nicht aus, daß letztere nicht auch wirksam sein könnte. Im Gegensatz zu den SKINNERSchen Tierversuchen, bei denen das Versuchstier nicht in der Lage war, einen verstärkenden Reiz auf ein Wirkverhalten zu beziehen, wenn zwischen diesem und dem verstärkenden Reiz weitere Wirkreaktionen lagen, scheint es für den humanpsychologischen Bereich nicht unbedingt gültig zu sein, daß ein verstärkender Reiz immer nur die Auftrittswahrscheinlichkeit der zeitlich unmittelbar vorausgehenden Wirkreaktion erhöht. Da die Schüler nicht - wie Tauben - "blind" lernen, sondern i.allg. in der Lage sind, den Zusammenhang zwischen einer vom Lehrer bestätigten Lösung (verstärkender Reiz) und der auf diese Lösung bezogenen Eigentätigkeit (Wirkverhalten) zu erkennen, wird einmal die zeitliche Wirksamkeit der vom Lehrer ausgehenden verstärkenden

Reize länger anhalten als dies bei Tierversuchen üblicherweise erwartet werden kann¹⁾, zum anderen werden auch Schüler, welche eine Lösung nur gedacht haben, verstärkt, da auch sie den Zusammenhang zwischen ihrer gedanklichen Aktivität und den vom Lehrer ausgehenden verstärkenden Reizen erkennen können²⁾. Sollten unsere Annahmen richtig sein, könnte bezüglich der Zahl der pro Schüler entfallenden Verstärkungen nicht mehr so ohne weiteres von einer Überlegenheit der Programmierten Unterweisung gegenüber unserer Form des herkömmlichen Unterrichts ausgegangen werden, da zwar - gemessen am Programm - relativ wenige individuelle und unmittelbare Verstärkungen im Sinne SKINNERS pro Schüler entfallen, insgesamt aber - bei Zugrundelegung der kollektiven Wirksamkeit eines verstärkenden Reizes und dessen Wirksamkeit auch im Falle zeitlicher Verzögerung - jedoch pro Schüler eine beträchtliche Anzahl von Verstärkungen entfallen³⁾. An dieser Stelle muß allerdings ausdrücklich vermerkt werden, daß Voraussetzung für das Wirksamwerden eines verstärkenden Reizes in jedem Fall eine innere Schüleraktivität und das rechtzeitige Finden der richtigen Lösung (vor Bekanntgabe derselben durch den Lehrer) ist. Wir werden im Zusammenhang mit dem individuellen Lerntempo darauf noch zu sprechen kommen. Zusammenfassend kann festgehalten werden, daß eine Überlegenheit der Programmierten Unterweisung gegenüber dem herkömmlichen Unterricht nur beschränkt - wenn überhaupt - auf Anzahl und Enge der zeitlichen Verbindung von verstärkendem Reiz und Wirkreaktion zurückgeführt werden kann.

- 1) In diesem Sinne kann der oben erwähnte Versuch von AUBLE und MECH interpretiert werden. Als Beleg dafür kann auch das von THORNDIKE untersuchte Phänomen der "Effektverbreitung" herangezogen werden, demzufolge eine Belohnung auch vorausgegangene und nachfolgende Reaktionen verstärkt. Weiter muß darauf hingewiesen werden, daß der aufgrund tierpsychologischer Versuche wiederholt erhobene Befund, daß bei Tieren die Wirksamkeit einer Verstärkung über einen längeren Zeitraum erhalten bleibt, sofern sie nur in der Verzögerungszeit ihre Aufgabenhaltung beibehalten, mit hoher Wahrscheinlichkeit für den humanpsychologischen Bereich in verstärktem Maße gilt, da die Schüler zum einen in der Lage sind, den Zusammenhang zwischen einer vom Lehrer bestätigten Lösung und ihrem Eigenverhalten auch über einen längeren Zeitraum hinweg herzustellen und zum anderen zunächst angesetzt werden muß, daß Schüler im Verlauf einer Unterrichtsstunde eine intensivere und länger anhaltende kontinuierliche Aufgabenhaltung beibehalten, als dies in tierpsychologischen Versuchen der Fall ist. Die Wirksamkeit von zeitlich verzögerten Verstärkungen kann auch mit Hilfe des Begriffs der "Erfolgserwartung" aufgezeigt werden. HULL ging davon aus, daß die einzelnen Unterschritte, welche ein komplexes Endverhalten aufbauen, mit dem verstärkten Endverhalten verbunden sind, so daß die Verstärkung des Endverhaltens auch die sie vorbereitenden Teilschritte betrifft (vgl. W. CORRELL: "Pädagogische Verhaltenspsychologie", München-Basel 1967, S. 80).
- 2) Dies gilt natürlich nur unter der Voraussetzung, daß "verdecktes" Verhalten überhaupt verstärkt werden kann.
- 3) Im günstigen Fall kann diese Anzahl bei konzentriert arbeitenden Schülern sogar größer sein als bei Verwendung eines Unterrichtsprogramms. Entsprechende Hinweise ergeben sich aus unseren Aufzeichnungen des Verstärkerverhaltens der an unserer Untersuchung beteiligten Lehrer.

- b) Ein Unterrichtsprogramm vom "Typ SKINNER" kann für sich in Anspruch nehmen, ein gewünschtes Schülerverhalten in verschiedenem Kontext zu verstärken und dieses Verhalten durch Übungseinheiten weiter zu verfestigen (vgl. 4.11 b), 4.11 e), 4.20 b) und 4.20 e)). Eine Analyse der von uns gewählten Form des herkömmlichen Unterrichts zeigt, daß aufgrund der relativ günstigen Aktivitätsverteilung zwischen Lehrer und Schüler (vgl. 4.31 a)) seitens des Lehrers genügend Möglichkeit besteht, ein Problem von verschiedenen Aspekten her anzugehen und die Schüler mittels Fragen und Impulsen zu entsprechenden Reaktionen aufzufordern. Gleichzeitig bietet unser Versuchsunterricht den Schülern ausreichend Gelegenheit, auf die entsprechenden Lehrerfragen und Impulse zu reagieren und verstärkt zu werden (vgl. 4.31 e) und 4.31 g)). Wenn hier (und unter Punkt a) dieses Kapitels) der Programmierten Unterweisung gegenüber dem herkömmlichen Unterricht keine oder nur geringe Vorteile eingeräumt werden, so bezieht sich das ausschließlich auf die von uns gewählte Form des Lehrgesprächs. Je größer der Aktivitätsüberhang seitens des Lehrers wird (z.B. durch dozierende Darstellungsweise), desto vorteilhafter hebt sich die Programmierte Unterweisung ab.
- c) Wir haben unter 4.20 c) und 4.20 d) gesehen, daß Unterrichtsprogramme ein starres Verstärkerprogramm in sich schließen und auch keine Rückgriffsmöglichkeit auf primäre Verstärkersysteme besteht. Abgesehen von ganz seltenen Ausnahmen kann davon ausgegangen werden, daß so gut wie ausschließlich dadurch verstärkt wird, daß der Schüler Einsicht in die rich-

tige Lösung gewinnt¹⁾. Hier scheinen uns nun die Bedingungen unseres herkömmlichen Vergleichsunterrichts vorteilhafter zu sein. Obwohl die Lehrer der Aufforderung unterlagen, auf jede an sie gerichtete Schüleraktivität zu reagieren (vgl. 4.31 e)), bestand diese Reaktion nicht immer in Form einer Verstärkung, häufig reagierte der Lehrer auch mit zusätzlichen Fragen, einem weiterführenden Impuls (vgl. 4.31 c)) u.a., so daß die Schüler nicht damit rechnen konnten, nach jedem Wirkverhalten verstärkt zu werden. Bei unserem herkömmlichen Vergleichsunterricht kann daher von einem variablen Verstärkersystem ausgegangen werden. Darüber hinaus stehen dem Lehrer neben der Bekanntgabe bzw. Bestätigung der richtigen Lösung noch eine Anzahl anderer Verstärker zur Verfügung, so z.B. soziale und individuelle Anerkennung, Notengebung etc. Gerade hier kann der herkömmliche Unterricht, sofern er nur von den verschiedenen Möglichkeiten der Verstärkung Gebrauch macht, von besseren Ausgangsbedingungen aus-

1) In diesem Zusammenhang schreibt CORRELL: "Hierin liegt eine gewisse Schwäche des gegenwärtigen Programmierens: Seine Wirksamkeit könnte beträchtlich gesteigert werden, wenn man neue und mehrere Verstärker abwechselnd verwenden könnte! ... Der Umstand, daß manche Kinder einzelne Programme als "langweilig" empfinden, hängt im wesentlichen nicht so sehr mit dem dargebotenen Stoff selbst zusammen, sondern mit der Eintönigkeit der Verstärkungen" (W. CORRELL, a.a.O., S. 260 f.).

gehen als die Programmierte Unterweisung¹⁾. Von Bedeutung wird dieser Vorteil insbesondere im Rahmen der Grundschulausbildung, da hier die Beschränkung auf das sachliche Verstärkersystem, wie es beim Programm vorliegt, nur bedingt und relativ kurzzeitig möglich ist²⁾ und zum anderen beim längerfristigen Einsatz von Unterrichtsprogrammen, da auch Schüler älteren Jahrgangs vermutlich eine Abwechslung in der Qualität der Verstärker benötigen³⁾. Wir müssen an dieser Stelle noch hinzufügen, daß SKINNER der Frage nach Qualität und Intensität der Verstärker nur untergeordnete Bedeutung zuerkennt, entscheidend ist nach seinem Dafürhalten, daß überhaupt verstärkende Reize vorhan-

-
- 1) Der Vorteil von unterbrochenen und variierenden Verstärkerverfahren wird auch von CORRELL betont. "Allgemein hat sich ergeben, daß alle unterbrochenen Verstärkungsverfahren insgesamt ein sehr viel stabileres Verhalten erzeugen als kontinuierliche Verstärkung. ... Ferner hat sich allgemein gezeigt, daß die variierenden Verstärkungsverfahren wiederum stabilere Verhaltensformen erzeugen als die nicht variierenden ... Verfahren. Außerdem haben ja die variierenden Verfahren noch den bereits erwähnten Vorteil, daß sie den Lorbeereffekt unterdrücken und ein relativ gleichmäßiges Verhalten entstehen lassen" (W. CORRELL, a.a.O., S. 90).
 - 2) Vgl. L. SCHENK-DANZIGER: "Schulreife und Erziehungsprobleme im Rahmen des Anfangsunterrichts", in: H. HETZER (Hrsg.): "Pädagogische Psychologie", Hb. d. Psych., Bd. 10, Göttingen 1959, S. 242
 - 3) Vgl. M. ZILLIG: "Psychologische Probleme der Oberstufe der Volksschule", in: H. HETZER (Hrsg.): "Pädagogische Psychologie", Hb. d. Psych., Bd. 10, Göttingen 1959, S. 284, und P.G. BRAUNE, a.a.O., S. 227

den sind¹⁾.

- d) Wir haben oben gesehen, daß bei der Programmierten Unterweisung jede negative Verstärkung unterbleibt. Dieser Vorteil gilt nicht für unseren herkömmlichen Unterricht. Obwohl keines der unter 4.31 geschilderten Unterrichtsmerkmale beinhaltet, daß der Lehrer mit negativen Verstärkermechanismen arbeiten muß, findet sich in der Praxis unseres herkömmlichen Unterrichts ein sehr häufiger Rückgriff auf negative Verstärker. Grundsätzlich hätte zwar der Lehrer die Möglichkeit, anstelle von negativ verstärkendem Verhalten überhaupt nicht zu reagieren bzw. durch entsprechende Zusatzfragen oder Impulse das gewünschte Schülerverhalten nachträglich herbeizuführen. Trotz dieser Möglichkeiten finden sich gehäuft Lehrerreaktionen, die im Falle einer fehlerhaften Schülerantwort von "das stimmt nicht" bis zu massiven Äußerungen wie "das ist ja unmöglich" und "stell Dich nicht so blöd an" reichen. Selbst ein relativ neutraler Lehrerkommentar wie z.B. "das stimmt nicht" hat zur Folge, daß der Schüler vor der gesamten Klasse bloßgestellt wird und dadurch jene Atmosphäre entstehen kann, die SKINNER in seinem Aufsatz "Why teachers fail" geschildert hat. Angesichts dieses Sachverhalts muß vermerkt werden, daß sich hier die Programmierte Unterweisung als eindeutig überlegen erweist. Da weiter - im Gegensatz zur Programmierten Unterweisung - aufgrund der

1) Vgl. B.F. SKINNER: "Die Wissenschaft vom Lernen und die Kunst des Lehrens", in: W. CORRELL (Hrsg.): "Programmiertes Lernen und Lehrmaschinen", Braunschweig 1966, S. 78

Fragetechnik der Lehrer (vgl. 4.31 j)) relativ häufig die Situation auftritt, daß Schüler falsche oder unvollständige Antworten geben, häufen sich in deren Folge auch die negativen Verstärker seitens des Lehrers. Bei unserer Form des herkömmlichen Unterrichts wird die Wahrscheinlichkeit des Auftretens ausdrücklicher negativer Verstärkungen darüber hinaus dadurch erhöht, daß alle ausdrücklich geäußerten Schüleraktivitäten über den Lehrer laufen (vgl. 4.31 b)e)f) und 4.31 h)) und im Falle falscher Schülerantworten hier negative Verstärkungen geradezu provoziert werden¹⁾. Allerdings muß unter Hinweis auf unsere experimentelle Fragestellung vermerkt werden, daß das häufige Vorkommen von negativen Verstärkern die Lehreffektivität nur indirekt beeinflußt²⁾ und zwar in jener atmosphärischen Weise, wie sie von SKINNER geschildert worden ist³⁾.

Bezüglich des Prinzips der Verstärkung kann abschließend festgestellt werden, daß allein von hier aus - im Gegensatz zu der Argumentation von SKINNER und CORRELL⁴⁾ - eine höhere Lehreffektivität

-
- 1) Dabei muß zusätzlich vermerkt werden, daß es sich hierbei häufig um emotional getönte negative Verstärker handelt (reinforcement), welche die atmosphärische Grundeinstellung der Schüler viel nachdrücklicher negativ beeinflussen als dies bei einfacher Nichtbestätigung (confirmation) der Fall ist. Vgl. dazu: H. BRAUNE, a.a.O., S. 125 ff.
 - 2) Vgl. E.R. HILGARD und G.H. BOWER: "Theorien des Lernens I", Stuttgart 1970, S. 162 f.
 - 3) Vgl. B.F. SKINNER: "Warum Lehrer versagen", in: W. CORRELL (Hrsg.): "Zur Theorie und Praxis des Programmierten Lernens", Darmstadt 1969, S. 93 ff.
 - 4) Repräsentativ für die Überbewertung der Rolle der Verstärkung bei der Programmierten Unterweisung ist die Äußerung von CORRELL: "Die Wirksamkeit des programmierten Lernens - gleichgültig welcher Art - beruht in erster Linie auf der größeren Verstärkung, die durch diese Form des Lernens vermittelt wird" (W. CORRELL, a.a.O., S. 259).

tät der Programmierten Unterweisung nicht ausreichend begründet werden kann. Gewisse Vorteile dieser Unterrichtsmethode gegenüber unserem herkömmlichen Unterricht können zwar im Hinblick auf die durchschnittliche Zahl der pro Schüler entfallenden Verstärkungen sowie deren zeitlichem Abstand vom jeweiligen Wirkverhalten vermutet werden, gleichzeitig muß jedoch einschränkend in Rechnung gestellt werden, daß eine direkte Übertragbarkeit der Funktionsprinzipien der Verstärkung im SKINNERschen Sinn auf die von uns untersuchten Lernprozesse zumindest fraglich ist. Das gehäufte Auftreten von negativen Verstärkern kann, wie wir oben gesehen haben, die Lehreffektivität des herkömmlichen Unterrichts nur indirekt negativ beeinflussen. Umgekehrt hat sich unser herkömmlicher Vergleichsunterricht hinsichtlich der Variabilität und Qualität des Verstärkerprogramms als überlegen erwiesen, wenn gleich letzterem von SKINNER keine wesentliche Bedeutung für den Lernprozeß zugemessen wird. Im übrigen können die unter 2.24 aufgeführten experimentellen Befunde¹⁾ zur Unterstützung unserer Ansicht indirekt mit herangezogen werden²⁾.

1) Vgl. auch A.A. LUMSDAINE: "Probleme beim Programmierten Unterricht", in: W. CORRELL (Hrsg.): "Zur Theorie und Praxis des Programmierten Lernens", Darmstadt 1969, S. 75 f.; D. ANGELL und A.A. LUMSDAINE: "The Effects of Prompting Trials and Partial-Correction Procedures on Learning by Anticipation", San Mateo, California: American Institute for Research, 1961

2) Eine direkte Verwertung dieser Befunde ist insofern problematisch, als diese Arbeiten fast durchweg so angelegt sind, daß eine Vergleichbarkeit nicht möglich ist. Solange keine Wiederholungsstudien und/oder Kreuzvalidierungen vorliegen, sollten diese Arbeiten im Kontext einer theoretischen Diskussion nicht überbewertet werden. Auf der anderen Seite kann - wenn auch mit Vorsicht - vermutet werden, daß im Falle des Zutreffens der SKINNERschen Argumentation nicht gerade in diesem Forschungsbereich die nichtsignifikanten Ergebnisse besonders häufig wären.

5.20 Das Prinzip der "aktiven Antwortformulierung"

Da für SKINNER das Wirkverhalten wesentlicher Bestandteil des Lernprozesses ist (vgl. 4.12), wird bei der Programmierten Unterweisung nach jedem Lernschritt von jedem einzelnen Schüler ein ausdrückliches, aktives Antwortverhalten verlangt (vgl. 4.20 h)). Auch hier kann grundsätzlich vermerkt werden, daß diese Bedingung von unserem herkömmlichen Unterricht nicht erfüllt wird. Zwar begünstigt unsere Form des herkömmlichen Unterrichts - verglichen mit anderen - die Möglichkeit der Schüler, ausdrücklich aktiv (vgl. 4.31 a)g) und 4.31 i)) d.h. verbal zu reagieren, da jede Schülerreaktion jedoch über den Lehrer zu laufen hat, sind die Schüler bei der Möglichkeit zur verbalen Aktivität auf den Lehrer angewiesen (vgl. 4.31 b)f)). Dies aber hat zur Folge, daß ein Schüler im Verlauf einer Unterrichtsstunde nur in beschränkter Zahl zu solchen verbalen Reaktionen gelangen kann, da der Lehrer bemüht sein wird, möglichst alle Schüler sukzessiv am Unterrichtsgespräch zu beteiligen. Selbst im Fall eines herkömmlichen Unterrichts mit relativ gleichmäßiger Aktivitätsverteilung (vgl. 4.31 a)) ist also die Möglichkeit zu verbaler, stoffbezogener, individueller Schülertätigkeit stark eingeschränkt¹⁾. Das Prinzip der ausdrücklichen verbalen und/oder motori-

1) Vgl. W. HOCHHEIMER: "Psychologische Einführung in die Thematik von Lehrmaschinen und Programmierem Unterricht", in: A. BIANCHERI u.a. (Hrsg.): "Programmierter Unterricht und Lehrmaschinen. Bericht Internationale Konferenz Berlin 1963", Berlin 1964, S. 251

schen Schüleraktivität wird auch von der herkömmlichen Didaktik befürwortet, es besteht ein deutlicher Trend innerhalb der gesamten Unterrichtslehre, vom dozierenden Unterricht weg zu Unterrichtsformen zu gelangen, welche dem Schüler möglichst viel aktives Handeln erlauben¹⁾. Eine weitere Kritik SKINNERs am herkömmlichen Unterricht im Hinblick auf die Schüleraktivität besteht darin, daß aufgrund der aversiven Lehrerkontrolle das Verhalten der Schüler häufig nicht stoffbezogen sei²⁾. Bisherige Erfahrungsberichte über den Einsatz programmierten Lernmaterials zeigen tatsächlich, daß die Schüleraktivität intensiver und gleichmäßiger stoffbezogen ist, daß die Schüler also konzentrierter arbeiten, als dies beim herkömmlichen Unterricht der Fall ist³⁾. Von hier aus sind auch die

-
- 1) Vgl. K. STÜCKER, a.a.O., S. 82 ff. und H. AEBLI: "Psychologische Didaktik", Stuttgart 1968, S. 119 f.
 - 2) In diesem Zusammenhang schreibt SKINNER: "Ein Schüler kann zwar anwesend sein, den Lehrer oder das Buch ansehen und trotzdem nicht aufpassen. Er ist hysterisch taub; seine Gedanken wandern; er ist ein Tagträumer. 'Geistige Müdigkeit' ist meistens kein Erschöpfungszustand, sondern ein nicht kontrollierbares Fluchtverhalten ... Der Stundenplan mißt mehr die Grenzen erzwungener Aufmerksamkeitskontrolle als die Länge der natürlichen Aufmerksamkeitsspanne". (B.F. SKINNER, a.a.O., S. 95)
 - 3) So konnte LINDVALL nachweisen, daß die Prozentwerte der Unaufmerksamkeit bei nichtprogrammiertem Lernen signifikant höher liegen als bei programmiertem Lernen. Diese Untersuchungen wurden im 1. und 4. Schuljahrgang durchgeführt, zur Messung der Unaufmerksamkeit wurde ein speziell für diese Untersuchungen entworfenes Verfahren herangezogen (vgl. C.M. LINDVALL: "Pupil attention and attitude under conditions of programmed instruction", Univ. of Pittsburgh 1964 (AERA-Paper)).

häufig geäußerten Hinweise zu verstehen, daß die Programmierete Unterweisung die Schüler mehr anstrengt als andere Unterrichtsformen¹⁾. Im Zusammenhang mit dem Prinzip der "overt response" muß allerdings die kritische Frage gestellt werden, inwieweit rein gedankliche Aktivität, sofern sie stoffbezogen ist, verbal formuliertem Verhalten, d.h. also dem Aussprechen des Gedachten bezüglich der Lehreffektivität, unterlegen ist. Auch hier scheint uns die radikale Forderung SKINNERS nach ausdrücklichem Antwortverhalten²⁾ von seinem experimentellen Vorgehen und seiner methodischen Konzeption her verstanden werden zu müssen. Kennzeichnend ist in diesem Zusammenhang die Äußerung von CORRELL, dem im deutschsprachigen Raum entschiedensten Verfechter der SKINNERschen Programmkonzeption: "Verstärkt kann indessen nur eine Verhaltensform werden, die tatsächlich geäußert wird. Dies bedeutet, der Schüler muß selbst aktiv werden, er muß sogar e i n d e u t i g aktiv werden, damit durch entsprechende Vergleiche festgestellt werden kann, ob die betreffende Reaktion im Sinne der Fragestellung richtig ist oder nicht"³⁾. Es ist kein Zufall, daß CORRELL nicht auf den Gedanken kommt, der Schüler könne nach vollzogener innerer Aktivität selbstständig überprüfen, ob seine gedachte Lösung mit der vorgegebenen Lösung übereinstimmt. Dahinter steht näm-

-
- 1) Vgl. H. HEINRICHS: "Roboter vor der Schultür?", Bochum 1964, S. 110 f.
 - 2) Vgl. B.F. SKINNER: "Lehrmaschinen", in: W. CORRELL (Hrsg.): "Programmiertes Lernen und Lehrmaschinen", Braunschweig 1966, S. 41
 - 3) W. CORRELL: "Pädagogische Verhaltenspsychologie", München-Basel 1967, S. 259

lich der SKINNERsche Denkansatz, demzufolge Wirkverhalten nur gelernt wird, wenn es einen verstärkenden Reiz hervorruft. Das operante Verhalten ist also die *Ursache* für das Auftreten eines verstärkenden Reizes, eine entsprechende Einwirkung auf die Umwelt ist bei "nur innerer Aktivität" nicht denkbar, folglich kann hier "nicht gelernt" werden¹⁾. Die unmittelbare Übertragung des in einer bestimmten experimentellen Situation gewonnenen Lernbegriffs auf schulisches Lernen zeigt sich ganz besonders deutlich in folgender Äußerung von CORRELL: "Geht es darum, eine neue Information zu vermitteln, so muß gewährleistet sein, daß der Lernende tatsächlich die Informations-elemente konstruiert, um anschließend dafür bestätigt (verstärkt) zu werden. Nur dasjenige, was verstärkt wird, wird auch gelernt"²⁾. Angesichts des nicht zu leugnenden Sachverhalts, daß auch bei nicht ausdrücklich geäußelter verbaler Aktivität gelernt wird, greift SKINNER auf den Begriff des "verdeckten Verhaltens" zurück (ein "Vorgänger" bzw. "Vorläufer" des operanten Verhaltens) und konstatiert, daß von einem solchen ausgegangen werden könne, "wenn die Verstärkung entweder automatisch erfolgt oder durch das nachfolgende Verhalten bewirkt wird"³⁾. Trotz dieses Zugeständnisses hat

-
- 1) SKINNER scheut sich selbst, entsprechend radikale Konsequenzen zu ziehen. Dies wird insbesondere dort deutlich, wo er sich mit der Problematik des menschlichen Denkens beschäftigt (vgl. B.F. SKINNER: "Verhaltenspsychologische Analyse des Denkprozesses", in: B.F. SKINNER und W. CORRELL: "Denken und Lernen", Braunschweig 1967, S. 26 ff.
 - 2) W. CORRELL: "Die Information im Unterrichtsprogramm", in: B.F. SKINNER und W. CORRELL: "Denken und Lernen", Braunschweig 1967, S. 73
 - 3) B.F. SKINNER, a.a.O., S. 27

SKINNER nie darauf hingewiesen, daß im herkömmlichen Unterricht eben die Bekanntgabe der Lösung durch den Lehrer für die "nur gedachte" Schüleraktivität jene automatische Verstärkung darstellt, welche wir oben zitiert haben. Wir vertreten hier die Ansicht, daß der Prozeß der Aufgabenlösung, also das die Aufgabenlösung bewirkende "Verhalten", subvokal erfolgt, die ausdrückliche Reaktion ist für den Lernprozeß nicht entscheidend, sie dient lediglich der Verfestigung der gefundenen Lösung und der Erhaltung der Aufmerksamkeitszuwendung des Schülers. Weiter wird von SKINNER die Konstruktionsantwortmethode im Hinblick darauf gefordert, daß vom Schüler seitens der Schule meistens verbale Reaktionen verlangt würden. Da die Auswahlantwortmethode dem Schüler jedoch lediglich beibringe, Reizkonstellationen zu unterscheiden, müsse hier auf die Konstruktionsantwortmethode zurückgegriffen werden, da, um eine verbale Reaktion zu lernen, dieselbe verstärkt werden müsse¹⁾. Hier muß kritisch angefügt werden, daß der Schüler nach unserem Dafürhalten nicht die verbale Reaktion lernen soll, sondern jene gedanklichen Aktivitäten, welche Voraussetzung für eine Aufgabenlösung und damit für eine verbale Äußerung sind. Damit soll nicht abgestritten werden, daß der Zwang zur aktiven Formulierung bereits während des Lernprozesses spätere verbale Äußerungen erleichtert, in keinem Fall kann jedoch der ausdrücklichen Antwort im Rahmen schulischer Lernprozesse jener Stellenwert zugemessen werden, wie dies seitens SKINNER und seiner Nachfolger geschieht. Wenn die herkömmliche Didaktik, wie wir oben gesehen haben, anstrebt, daß

1) Vgl. hierzu auch W. CORRELL, a.a.O., S. 73 f.

die Schüler möglichst häufig aktiv verbal und motorisch reagieren sollen, so vor allem deswegen, weil damit die Stoffzuwendung und damit letztlich die gedankliche Aktivität erhöht wird, keinesfalls aber, weil die gedankliche Aktivität geringer eingeschätzt wird¹⁾. Es kann an dieser Stelle nicht geklärt werden, inwieweit der Zwang zur ausdrücklichen Reaktion über die Verfestigung der gefundenen Lösungen und der Erhaltung einer kontinuierlichen und erhöhten Aufmerksamkeitsspannung eine höhere Lehreffektivität bewirkt als dies der herkömmliche Unterricht zu leisten vermag. Sicher ist, daß auch ohne aktive Reaktion gelernt werden kann. Im übrigen bietet die von uns gewählte Form des herkömmlichen Unterrichts den Schülern weitaus mehr Gelegenheit ausdrücklich zu reagieren als die darbietend-dozierende Form, da nicht nur, wie wir bereits oben erwähnt haben, das Verhältnis von Lehrer- und Schüleraktivität relativ ausgewogen ist, sondern darüber hinaus auch nicht aufgerufene Schüler ständig damit rechnen müssen, vom Lehrer zu einer verbalen Aktivität aufgefordert zu werden (vgl. 4.31 h) und 4.31 i)²⁾.

Bei Berücksichtigung aller angeführten Aspekte kann nach unserem Dafürhalten auch das Prinzip der aktiven

1) Vgl. K. STÖCKER, a.a.O., S. 82 ff.

2) Dies könnte allerdings auch als aversive Lehrerkontrolle mit all ihren negativen Auswirkungen interpretiert werden. Zumindest für Schüler mit ablehnender Einstellung gegenüber Schule und Lehrer muß ein solcher Mechanismus angesetzt werden.

Antwortformulierung nicht herangezogen werden, um eine höhere Lehreffektivität der Programmierten Unterweisung schlüssig zu belegen. Auch hier liegen experimentelle Befunde vor (vgl. 2.23), welche, analog zu 5.10, unseren theoretisch abgeleiteten Befund indirekt bestätigen¹⁾.

5.30 Das Prinzip des "shaping of behavior"

Das den SKINNER-Programmen zugrunde gelegte Prinzip des shaping besteht darin, das mehr oder weniger komplexe endgültige Lernziel "in kleinste Erkenntnisschritte aufzugliedern, die in Form von einfachen Feststellungen und Fragen dargeboten werden. Jede Antwort ... wird sofort bewertet, d.h. im positiven Fall wird die betreffende Verhaltensform sofort verstärkt"²⁾. SKINNER selbst sagt dazu: "Wenn wir einmal die besondere Art einer Folgeerscheinung, die wir Verstärkung nennen, hergestellt haben, erlauben es unsere Methoden, das Verhalten eines Organismus fast nach Belieben zu formen"³⁾.

1) Vgl. L.J. ISSING, a.a.O., S. 107; A.P. CSANYI, R. GLASER, J.H. REYNOLDS: "Programmierungsmethode und Antwortart bei einem Experiment zur Schulung der Aussprache mit visuellen Hilfen", in: W. CORRELL: "Zur Theorie und Praxis des Programmierten Lernens", Darmstadt 1969; E.B. FRY: "A Study of Teaching Machines Response Modes", in: A.A. LUMSDAINE und R. GLASER (Hrsg.): "Teaching Machines and Programed Learning", Washington 1960

2) W. CORRELL: "Programmiertes Lernen und schöpferisches Denken", München 1965, S. 11

3) B.F. SKINNER: "Die Wissenschaft vom Lehren und die Kunst des Lernens", in: W. CORRELL (Hrsg.): "Programmiertes Lernen und Lehrmaschinen", Braunschweig 1966, S. 67

Wenn wir unter 5.10 von Verstärkung gesprochen haben, dann ließen wir dort den Aspekt des Verhaltensrepertoires (Programms), welches verstärkt werden soll, außer acht. Dieses systematisch aufgebaute Verhaltensprogramm ist nun aber tatsächlich ein außerordentlich wesentliches Konstruktionsprinzip der programmierten Unterweisung. Unter enger Anlehnung an das Prinzip der operationalen Verhaltens- bzw. Lernzielbeschreibung wird das angestrebte Endverhalten sorgfältig analysiert und systematisch in aufeinander aufbauende Verhaltenseinheiten gegliedert¹⁾. In der Tat geht in den seltensten Fällen dem herkömmlichen Unterricht eine so sorgfältige Analyse der Lernziele und eine dermaßen darauf basierende Unterrichtsstrategie voraus. Selbst wenn das endgültige Lernziel und wesentliche Zwischenschritte hinlänglich operational formuliert sind, bleiben Inhalt und Sequenz der zur Erreichung des Lernziels notwendigen kleinen Lernschritte i.allg. dem Zufall überlassen, jedenfalls aber werden sie vorher nicht so systematisch geplant. Selbst bei strenger Lehrerkontrolle aller unterrichtlichen Aktivitäten (vgl. 4.31 b) und 4.31 h)) wechseln Phasen eines zähen Verharrens bei einer Übungsaufgabe, d.h. eine Folge kleinster Lernschritte, mit Phasen übereilter und zu schwieriger Lernschritte (vgl. 4.31 j)). So kann z.B. ein schwacher Schüler bei der Lösung einer Aufgabe systematisch nach Absolvierung kleinster Ver-

1) "Im Sinne dieser Charakteristika könnte man ein Programm definieren als ein Instrument zur schrittweisen Veränderung des Verhaltens durch Herausforderung entsprechender, genau vorausgeplanter Reaktionen und ihre sofortige Verstärkung, so daß der Lernende von einem Ausgangsverhalten zu einem präzise festgestellten Zielverhalten fortschreitet" (W. CORRELL: "Pädagogische Verhaltenspsychologie", München-Basel 1967, S. 250).

haltenseinheiten, wie sie nicht einmal im diesbezüglich streng konzipierten SKINNER-Programm zu finden sind, verstärkt werden¹⁾. Wenn aber wenig später eine Unterrichtsphase relativ schwieriger Lernschritte erfolgt, ist derselbe Schüler weder zu aktiven noch zu "verdeckten" Reaktionen in der Lage, welche der vom Lehrer angestrebten Lösung entsprechen. Daraus folgt, daß er in dieser Unterrichtsphase keinerlei Verstärkung erfährt, auch nicht in dem erweiterten Sinne, wie wir das oben geschildert haben. Da spätere Unterrichtsphasen aber i.allg. auf vorausgehenden aufbauen, wird die Wahrscheinlichkeit, daß insbesondere schwache Schüler noch Verstärkungen erfahren, immer geringer. Grund dafür ist, daß trotz der vom Lehrer häufig gegebenen Verstärkungen diese nur einen Teil - fast immer den besseren - der Schüler erreichen, da nur sie alle Unterrichtsphasen mit hinlänglichem Erfolg absolviert haben und somit jene aktiven und verdeckten Wirkreaktionen tätigen konnten, welche Voraussetzung für das Wirksamwerden der Verstärker sind. Was also den herkömmlichen Unterricht von der Programmierten Unterweisung deutlich unterscheidet, ist das Fehlen einer konsequent durchgeführten Lernschrittstrategie, welche jeden Schüler erreicht und somit über eine punktuelle Verstärkung während einzelner Unterrichtsphasen hinaus zu einem systematischen Ver-

1) Entsprechende Erfahrungen machten wir im Verlauf der Unterrichtsprotokollierung unserer Untersuchung.

stärkungsprogramm führt¹⁾. Nach unserer Ansicht läßt sich auch von hier aus das eigentümlich starre Verhalten schwacher Schüler bei Prüfungsaufgaben erklären, diese können lediglich die wenigen aufgrund der punktuellen Verstärkung gelernten Verhaltensweisen bei der Lösung von Prüfungsaufgaben anwenden und müssen in dem Moment scheitern, in dem eine Aufgabe auch nur geringfügig vom Konzept der eingeübten Verhaltensweisen abweicht. Abgesehen von der Möglichkeit gezielter, systematischer und kontinuierlicher Verstärkung bei der Programmierten Unterweisung kann davon ausgegangen werden, daß die intensiven Überlegungen bezüglich des Aufbaus einer Sequenz optimal angeordneter kleinster Lernschritte zu einem didaktischen Konzept führt, welches i.allg. besser ist als die entsprechenden Konzepte des herkömmlich vorbereiteten und gehaltenen Unterrichts. Abgesehen von der Frage, ob die besser durchdachte didaktische Struktur der Lehrprogramme oder aber die damit zusammenhängende Möglichkeit zur systematischen und kontinuierlichen Verstärkung die Lernprozesse nachhaltiger positiv beeinflusst, kann festgehalten werden, daß hier ein echter Vorteil der Programmierten Unterweisung gegenüber dem herkömmlichen Unterricht vor-

1) In diesem Zusammenhang schreibt SKINNER: "Es ist gewöhnlich nötig, das Verhalten in Reaktionsblöcken zu verstärken, wie beim Korrigieren eines Arbeitsbogens oder einer Heftseite. Die Reaktionen innerhalb eines solchen Blocks dürfen nicht aufeinander bezogen sein. Die Antwort auf die Frage darf nicht von der Antwort auf eine andere abhängen. Die Zahl der Stufen, durch die man sich einem komplexen Verhaltensmuster progressiv nähern kann, ist deshalb klein und die Aufgabe wird umso schwieriger. Selbst das modernste Arbeitsheft für den Anfangsrechnunterricht gibt nicht im entferntesten ein Beispiel für ein leistungsfähiges Programm. ..." (B.F. SKINNER, a.a.O., S. 74)

liegt, woraus sich die Erwartung einer höheren Lehr-effektivität ableiten läßt¹⁾. An dieser Stelle muß noch kurz auf zwei Einwände hingewiesen werden, welche sich gegen das Prinzip des "shaping of behavior" richten. Einmal wird dieser Form der Lehrstrategie vorgeworfen, daß sie zu "mechanisiertem" und "linearem" Denken verleite, d.h. daß sie nicht dazu in der Lage wäre, "wirkliches Denken" zu lehren und deswegen sogar für manche Unterrichtsbereiche geradezu verhängnisvoll sei. Ein engagierter Kritiker in diesem Sinne ist AEBLI²⁾. Wir haben bereits oben auf die Problematik dieser Argumentation hingewiesen, so daß wir uns hier weitere Ausführungen sparen können. Weiter wird von Kritikern angeführt, dieses lineare System sei unbeweglich, d.h. es könne sich den speziellen Bedürfnissen einzelner Schüler nicht anpassen³⁾.

-
- 1) In diesem Sinne kann LUMSDAINE interpretiert werden, wenn er schreibt: "Aber was die Aufmerksamkeit des Programmierers vor allem fesselt - und wohl zu Recht fesselt - ist sicherlich die geschickte Handhabung von Einhilfestichwörtern, nicht die Handhabung von Belohnungsmethoden" (A.A. LUMSDAINE, a.a.O., S. 77), wobei an dieser Stelle hinzugefügt werden muß, daß LUMSDAINE lerntheoretisch an GUTHRIE orientiert ist und von hierher auch seine Erklärungsprinzipien formuliert. Daß die Vorgabe der richtigen Einhilfen eine höhere Lehreffektivität bewirken kann, zeigt u.a. die Arbeit von L.M. STOLUROW and H. LIPPERT: "Prompting, Confirmation and Over-learning in the Automated Teaching of a Sight Vocabulary", in: J.P. De CECCO (Hrsg.): Educational Technology", New York 1964, S. 187-197 .
 - 2) Vgl. H. AEBLI: "Programmierte Ausbildung in didaktischer Sicht", Deutsche Schule, Hannover 58, 1966, S. 315-326
 - 3) Vgl. N.A. CROWDER: "Unterschiede zwischen linearem und verzweigtem Programmieren", in: W. CORRELL (Hrsg.): "Zur Theorie und Praxis des Programmierten Lernens", Darmstadt 1969, S. 152 u. 156

Bei konsequent konzipierten SKINNER-Programmen mit kleinsten Lernschritten tritt nach unserer Auffassung das Problem der fehlenden Anpassungseignung nicht auf, da die Schüler aufgrund der einfachen Lernschritte jederzeit wissen, was für Reaktionen von ihnen erwartet werden. Problematisch wird die Konzeption des linearen Programms, wenn die Lernschritte zwar linear aufgebaut sind aber nicht die Bedingung eines geringen Schwierigkeitsgrades erfüllen. Dies trifft tatsächlich bei sehr vielen linearen Programmen, insbesondere im deutschsprachigen Raum zu, da hier das lernpsychologische Konzept der Programmierten Unterweisung, wie es von SKINNER vertreten wird, stark in den Hintergrund getreten ist. Bei solchen Programmen kann es tatsächlich geschehen, daß ein Schüler an einer wesentlichen Stelle im Kontext der einzelnen Lernschritte eine entscheidende Aufgabe nicht zu lösen vermag. Da er keine Möglichkeit zur Rückfrage hat und das Programm selbst - im Gegensatz zum verzweigten Programmtyp, bei welchem eine solche Möglichkeit grundsätzlich besteht - auch nicht weiter hilft, kann er im weiteren Verlauf der Durcharbeitung des Programms mehr oder weniger nachdrücklich behindert sein. Von der Anlage her ist in dieser Hinsicht die von uns gewählte Form des herkömmlichen Unterrichts anpassungsfähiger. Der Schüler hat hier die Möglichkeit nachzufragen, darüber hinaus kann der Lehrer aus der Schülerreaktion auf seine Fragen hin erschließen, ob eine bestimmte Verhaltensform aufgebaut ist oder nicht und er kann dadurch seine weitere Lernstrategie jeweils anpassen. Die erste Möglichkeit, also das Nachfragen, wurde von Schülern - wie aus unseren Beobachtungsprotokollen hervorgeht - so gut wie nicht wahrgenommen. Die zweite Möglichkeit wird vom Lehrer bei dieser Form

des herkömmlichen Unterrichts zweifellos grundsätzlich genutzt, aber die Wirkung der Anpassung seitens des Lehrers kann als relativ gering veranschlagt werden, da sie angesichts der Vielzahl der Schüler wiederum nur punktuell und unsystematisch erfolgen kann¹⁾. Zudem haben solche Anpassungshandlungen oft zur Folge, daß viele Schüler von der eigentlichen Lernrichtung abgelenkt werden, da plötzlich unterrichtliche "Querlinien" auftreten, welche insbesondere diejenigen Schüler verwirren können, an welche die Anpassungshandlungen des Lehrers nicht gerichtet sind²⁾.

1) Vgl. L.N. LANDA, A.M. ORLOVA, G.G. GRANIK: "Prinzipien des Programmierten Unterrichts in russischer Sprache", in W. CORRELL (Hrsg.): "Zur Theorie und Praxis des Programmierten Lernens, Darmstadt 1969, S. 240 f. Die Autoren argumentieren von der Position der pädagogischen Kybernetik aus und stellen fest, daß das Rückkoppelungsprinzip im herkömmlichen Unterricht nur unzulänglich beachtet werde. Da ein intakter Regelkreis zwischen dem Lehrer und einem Schüler nur selten und zeitlich kurzfristig bestehe, gleiche der Lehrer einem Arzt, der eine Diagnose zu stellen habe, ohne daß er die Ursachen kenne. Die Folge davon sei ein häufiger Mißerfolg seiner pädagogischen Bemühung.

2) An dieser Stelle bietet sich das computergesteuerte Lehrprogramm als optimale Lernhilfe an, da hier zwischen Programm und Schüler ständig ein intakter Regelkreis besteht, aufgrund dessen dem Schüler ein individuelles Stoffangebot gemacht werden kann, ohne daß andere Schüler davon betroffen werden.

5.40 Das Prinzip der sozialen Entlastung der Lernprozesse

Es kann nicht bezweifelt werden, daß die Programmierte Unterweisung den Lernprozeß aus der Abhängigkeit von häufig emotional gestörten Lehrer-Schüler- und Schüler-Klasse-Beziehungen herauslöst und in einer sachlichen, emotional entspannten Atmosphäre beläßt. Dies ist vor allem für diejenigen Schüler von Bedeutung, deren persönliche Beziehungen insbesondere zum Lehrer aber auch deren Verhältnis zum Klassenverband gestört sind. Darunter fallen häufig auch Schüler mit einer hohen Angstbereitschaft, da sie, wie wir aus der Persönlichkeitspsychologie wissen, gehäuft störanfällige soziale Beziehungen aufweisen¹⁾. Die negativen Auswirkungen solcher emotionaler Störungen hat SKINNER, wie wir oben gesehen haben, eindrucksvoll aufgewiesen. Selbst wenn wir davon ausgehen, daß SKINNERs einschlägige Schilderung sich offensichtlich an einem besonders krassen Fall von autoritär geführtem Unterricht orientiert, bleibt der Kern der SKINNERschen Argumentation davon unberührt. Es ist kein lehrerbezogener Unterricht denkbar, der in der Lage wäre, die Lernprozesse von solchen sozialen Span-

1) Insbesondere die tiefenpsychologischen Persönlichkeitstheorien haben auf die Bedeutung der Angst für das menschliche (Sozial-) Verhalten hingewiesen (vgl. P.R. HOFSTÄTTER: "Tiefenpsychologische Persönlichkeitstheorien", in: H. LERSCH und H. THOMAE (Hrsg.): "Persönlichkeitsforschung und Persönlichkeitstheorie", Hb. d. Psych., Bd. 4, Göttingen 1960, S. 579 ff.

nungen freizuhalten¹⁾, wobei an dieser Stelle nicht verschwiegen werden soll, daß der Einbezug in einen sozialen Kontext für Lernprozesse auch förderlich sein kann (man denke dabei nur an diejenigen Schüler, welche "dem Lehrer zuliebe" arbeiten, ein Fall, der in der Schulpraxis nicht allzuselten vorkommt). Ungeachtet dieser Möglichkeit halten wir die Gefahren einer Emotionalisierung der Lernprozesse²⁾ - und hier befinden wir uns in voller Übereinstimmung mit SKINNER - für gravierender als deren mögliche Vorteile. Hier eröffnet die Programmierte Unterweisung tatsächlich positive Möglichkeiten der Unterrichtsgestaltung. Dies gilt umso mehr, als beim herkömmlichen Unterricht die Schüler hinsichtlich ihrer stoffbezogenen Aktivität auf den Lehrer angewiesen sind (vgl. 4.31 f)). Einschränkung muß allerdings gesagt werden, daß sich die Vorteile dieser "emotionalen Entkrampfung der Lernsituation" durch den Einsatz von Unterrichtsprogrammen möglicherweise erst bei länger dauernder Verwendung der Programmierten Unterweisung in einem Ausmaß bemerkbar machen dürften, daß gegenüber dem herkömmlichen Unterricht eine deutliche Leistungssteigerung erkennbar ist. Es kann sogar vermutet werden, daß negative Einstellungen dem Lehrer und der Schule gegenüber zunächst auch auf das Programm übertragen werden, da dieses von den betroffenen Schülern zunächst im Kon-

-
- 1) W. HOCHHEIMER, a.a.O., S. 250 schreibt in diesem Zusammenhang: "Weil zwischenmenschlicher Unterricht nie total sachlich erfolgen kann und so leicht subjektive Affekte erregt, gibt es psychologisch keinen 'reinen' Unterricht".
 - 2) Vgl. W. HOCHHEIMER: "Zur Tiefenpsychologie des pädagogischen Feldes", in: Die Deutsche Schule, 51. Jg., H. 3, März 1959, S. 101-119

text der als grundsätzlich bedrohlich oder feindlich erlebten schulischen Umwelt gesehen wird. Es ist daher fraglich, ob der von der Hypothese der "emotionalen Entkrampfung der Lernsituation" her bestimmte Einsatz von Unterrichtsprogrammen in entsprechend ausgewählten Schülergruppen kurzfristig die erwünschten Wirkungen zeigt.

5.50 Das Prinzip des individuellen Lerntempos.

Aufgrund des Konstruktionsprinzips der kleinen Lernschritte beim SKINNER-Programm hat auch der weniger intelligente und fachspezifisch weniger leistungsfähige Schüler die Möglichkeit, die gewünschten Verhaltensformen vollständig zu erlernen. Er wird die Lernschritte zwar unter größerem Zeitaufwand durchlaufen, dabei aber jeweils die richtige Lösung selbstständig finden und formulieren. Das Prinzip des individuellen Lerntempos ermöglicht es also auch weniger intelligenten und leistungsfähigen Schülern, die notwendigen Lernziele zu erreichen. Dasselbe gilt für Schüler, die unbesehen ihrer Intelligenz ein langsames Arbeitstempo aufweisen. Da jeder Schüler sein eigenes Programm hat, kann er die ihm gemäße Lerngeschwindigkeit selbst bestimmen und von Fall zu Fall variieren, ohne daß dadurch andere Schüler gestört werden. Völlig anders zeigt sich der Sachverhalt bei unserem herkömmlichen Unterricht. Hier hängt die Geschwindigkeit des Vorgehens in erster Linie vom Lehrer ab (vgl. 4.31 b)). Der Lehrer wird sich dabei i.allg. bezüglich des Tempos des unterrichtlichen Vor-

gehens nach einem von ihm taxierten "durchschnittlichen Schüler" richten, wobei die Kriterien zur Ermittlung eines "durchschnittlichen Schülers" oft sehr willkürlich gewählt sind. Da zudem auch die Größe der Lernschritte i.allg. nach diesem "durchschnittlichen Schüler" ausgerichtet ist, hat ein solches Vorgehen zur Folge, daß für den weniger intelligenten Schüler die Lernschritte häufig zu groß und das unterrichtliche Tempo meistens zu schnell ist. Er kommt also selten dazu, das vom Lehrer gewünschte Wirkverhalten, ob ausdrücklich oder verdeckt, zu vollziehen. Das aber hat zur Folge, daß es von den Lehrerverstärkungen nicht erreicht wird. Mit fortschreitendem Verlauf des Unterrichts werden für diesen Schüler die Lernbedingungen immer schlechter, da bei schulischen Lernstoffen i.allg. davon ausgegangen werden kann, daß die einzelnen Lernphasen aufeinander aufbauen, am Ende unterbleibt im krassen Fall jede stoffbezogene Aktivität, der Schüler resigniert. Der hier ausgeführte Gedankengang gilt entsprechend für den langsam arbeitenden Schüler. Weniger folgenschwer wirkt sich die Lehrerorientierung am "durchschnittlichen Schüler" für die intelligenten und schnell arbeitenden Schüler aus. Für sie besteht der herkömmliche Unterricht häufig im Leerlauf, da sie die jeweils erwünschte Verhaltensweise bereits zu einem Zeitpunkt erlernt haben, zu dem der Lehrer diese immer noch einübt. Da sie aber bezüglich des Angebots an weiterführender Information und entsprechender Fragen auf den Lehrer angewiesen sind (vgl. 4.31 b) und 4.31 f)), entstehen für sie mehr oder weniger lange Leerzeiten. Die Folgen sind Langeweile und eine negative emotionale Tönung, welche die Lernbereitschaft nachteilig beeinflusst. Es kann also festgehalten werden, daß der

herkömmliche Unterricht, sofern sich der Lehrer an die gesamte Klasse wendet, auf Unter- bzw. Überforderung bestimmter Schülergruppen von vornherein angelegt ist. Hier bietet die Programmierte Unterweisung die beste Möglichkeit, diesen Zustand zu beheben. Die Gewährleistung eines individuellen Lerntempos ist neben der systematischen und kontinuierlichen Verhaltensformung nach unserem Dafürhalten der entscheidende Vorteil der Programmierten Unterweisung gegenüber dem herkömmlichen Unterricht. Da die aus dem "mittleren Lerntempo" resultierenden negativen Bedingungen des herkömmlichen Unterrichts sich für die intelligenteren Schüler weniger folgeschwer auswirken als für die unintelligenten, kann erwartet werden, daß letztere in ihren Lernprozessen durch die Programmierte Unterweisung besonders begünstigt sind. Allerdings müssen wir darauf hinweisen, daß vor allem in Hinblick auf die Tatsache, daß die meisten linear angelegten Programme das Konzept der kleinen Lernschritte nur ungenügend verwirklichen, auch bei programmiertem Lernmaterial die Intelligenz i.allg. den zentralen Faktor für den Wissenserwerb darstellt¹⁾.

1) Vgl. H. SCHIEFELE: "Programmierte Unterweisung", München 1964, S. 95 f.

6.00 Hypothesen

Die Formulierung unserer Hypothesen basiert auf der unter Punkt 5.00 berichteten vergleichenden Betrachtung von Programmierter Unterweisung und der in unserer Untersuchung verwendeten Form des herkömmlichen Lehrerunterrichts sowie den unter Punkt 2.00 berichteten Forschungsergebnissen. Wir werden bei jeder Hypothese eigens auf diejenigen Punkte verweisen, auf welche sie in erster Linie zurückgeführt werden muß¹⁾. Angesichts dieser Verweise verzichten wir auf eine nochmalige Zusammenfassung der Gesichtspunkte, welche der jeweiligen Hypothese zugrunde liegen.

-
- 1) Diese Hinweise können lediglich auf Schwerpunkte aufmerksam machen. Insbesondere die Hypothesen, welche die höhere Lehreffektivität der Programmierten Unterweisung ohne Berücksichtigung weiterer adressatenbezogener Maßkriterien betreffen, basieren letztthin auf der gesamten vergleichenden Betrachtung von Programmierter Unterweisung und herkömmlichem Unterricht. Abgesehen von den deutlich erkennbaren Vorteilen dieser Methode, welche wir unter den Punkten 5.30 und 5.50 geschildert haben, muß davon ausgegangen werden, daß auch das Prinzip der sozialen Entlastung und - wenn auch sehr eingeschränkt - die Handhabung der Verstärkung und der Zwang zu ständiger aktiver Reaktion systematische Varianzanteile liefern, welche sich summieren und somit insgesamt zu der von uns erwarteten höheren Lehreffektivität der Programmierten Unterweisung führen können.

6.10 Hypothesen, welche die Ausprägung der Lernleistung betreffen

6.11 Lernzielkategorie "knowledge":

Der durchschnittliche Lernerfolg ist bei Programmierter Unterweisung (PU) signifikant höher als im herkömmlichen Unterricht (KU) (vgl. 2.10, 5.30 und 5.50).

$$H\ 1 : \quad M_{B/\text{knowl.}/PU} > M_{B/\text{knowl.}/KU}$$

$$H\ 2 : \quad M_{A/\text{knowl.}/PU} > M_{A/\text{knowl.}/KU}$$

$$H\ 3 : \quad M_{B/\text{knowl.}/PU} > M_{A/\text{knowl.}/KU}$$

$$H\ 4 : \quad M_{A/\text{knowl.}/PU} > M_{B/\text{knowl.}/KU}$$

6.12 Lernzielkategorie "comprehension":

Der durchschnittliche Lernerfolg ist bei Programmierter Unterweisung signifikant höher als im herkömmlichen Unterricht (vgl. 2.10, 5.30 und 5.50).

$$H\ 5 : \quad M_{B/\text{compr.}/PU} > M_{B/\text{compr.}/KU}$$

$$H\ 6 : \quad M_{A/\text{compr.}/PU} > M_{A/\text{compr.}/KU}$$

$$H\ 7 : \quad M_{B/\text{compr.}/PU} > M_{A/\text{compr.}/KU}$$

$$H\ 8 : \quad M_{A/\text{compr.}/PU} > M_{B/\text{compr.}/KU}$$

6.20 Hypothesen, welche den Zusammenhang von Intelligenz¹⁾ und Lernleistung betreffen

6.21 Lernzielkategorie "knowledge":

- H 9 : Der Zusammenhang von Intelligenz und Lernleistung ist bei beiden Unterrichtsmethoden signifikant (vgl. 2.30 und 5.50).
- H 10 : Der Zusammenhang von Intelligenz und Lernleistung ist im herkömmlichen Unterricht signifikant größer als bei Programmierter Unterweisung.
- H 11 : Die im Rahmen der Hypothesen 1 - 8 behauptete höhere Lehreffektivität der Programmierten Unterweisung wirkt sich bei Schülern mit "geringer Intelligenz"²⁾ in signifikant größerem Ausmaß aus als bei Schülern mit "hoher Intelligenz"³⁾ (vgl. 2.30 und 5.30).

6.22 Lernzielkategorie "comprehension":

- H 12 : Der Zusammenhang von Intelligenz und Lernleistung ist bei beiden Unterrichtsmethoden signifikant (vgl. 2.30 und 5.50).

1) Darunter verstehen wir den SW-Gesamtwert, welcher mit Hilfe des I-S-T erhoben wurde (vgl. 7.821).

2) Darunter verstehen wir Schüler, welche im unteren Fünftel unserer empirisch gefundenen Intelligenzverteilung liegen.

3) Darunter verstehen wir Schüler, welche im oberen Fünftel unserer empirisch gefundenen Intelligenzverteilung liegen.

- H 13 : Der Zusammenhang von Intelligenz und Lernleistung ist im herkömmlichen Unterricht signifikant größer als bei Programmierter Unterweisung.
- H 14 : Die im Rahmen der Hypothesen 1 - 8 behauptete höhere Lehreffektivität der Programmierten Unterweisung wirkt sich bei Schülern mit "geringer Intelligenz" in signifikant größerem Ausmaß aus als bei Schülern mit "hoher Intelligenz" (vgl. 2.30 und 5.30).

6.30 Positive Verstärkungen

- H 15 : Es ist *k e i n* signifikanter Zusammenhang zwischen der Häufigkeit an positiven Verstärkungen¹⁾ und der Lernleistung im herkömmlichen Unterricht nachweisbar (vgl. 2.24 und 5.10)²⁾.

-
- 1) Der hier verwendete Begriff der "positiven Verstärkung" deckt sich mit der Definition, welche im Rahmen der Besprechung unserer Beobachtungskategorien berichtet wird (vgl. 7.71: "Individuelle positive Verstärkungen").
- 2) Die Hypothesen 15, 16, 17 und 18 enthalten die Erwartung, daß zwischen den untersuchten Bedingungen *k e i n e* signifikanten Zusammenhänge bestehen. Dies ist eine ungewöhnliche Hypothesenformulierung, sie findet sich jedoch im Rahmen der Unterrichtsfor-schung recht häufig (vgl. P. STRITTMATTER: "Präsen-tationsmodi von Lehrprogrammen", Weinheim 1970, S. 185 ff.). In unserem Fall sind die Hypothesen die logische Konsequenz aus der Analyse der heute noch häufig vertretenen Meinung, daß die regelmäßige und unmittelbar verabreichte Verstärkung sowie die Kon-struktionsantwortmethode die wesentlichen Bedingungs-faktoren für die Überlegenheit der Programmierten Unterweisung gegenüber dem herkömmlichen Unterricht seien. Bezüglich der Aussagekraft solcher Hypothesen finden sich im Rahmen der Interpretation nähere Er-läuterungen.

- H 16 : Bei Schülern, welche im herkömmlichen Unterricht extrem selten positiv verstärkt werden¹⁾, ist kein signifikant höherer Lernzuwachs $X_{PU} - X_{KU}$ nachweisbar als bei Schülern, welche im herkömmlichen Unterricht extrem häufig positiv verstärkt werden²⁾ (vgl. 2.24 und 5.10).

6.40 "Unterrichtsbezogene Schüleraktivität"

- H 17 : Es ist kein signifikanter Zusammenhang zwischen der Häufigkeit "unterrichtsbezogener Schüleraktivität"³⁾ und der Lernleistung im herkömmlichen Unterricht feststellbar (vgl. 2.23 und 5.20).

-
- 1) Darunter verstehen wir alle Schüler, welche aufgrund der Daten unserer Beobachtungsprotokolle im unteren Fünftel der Häufigkeitsverteilung liegen.
 - 2) Darunter verstehen wir alle Schüler, welche aufgrund der Daten unserer Beobachtungsprotokolle im oberen Fünftel der Häufigkeitsverteilung liegen.
 - 3) Der hier verwendete Begriff deckt sich mit der Definition, welche im Rahmen der Diskussion unserer Beobachtungskategorien berichtet wird (vgl. 7.71: "Unterrichtsbezogene Schüleraktivität"). Es handelt sich also nicht um eine "Schülereigenschaft" sondern ausschließlich um die Summe aller seitens der Schüler protokollierten "unterrichtsbezogenen Aktivitäten". Dies ist sowohl für die Bestimmung des Skalierungsniveaus als auch für die Diskussion unserer Ergebnisse von Bedeutung.

H 18 : Bei Schülern, welche im herkömmlichen Unterricht extrem selten "unterrichtsbezogene Aktivität" aufweisen¹⁾, ist kein signifikant höherer Lernzuwachs ($\bar{X}_{PU} - \bar{X}_{KU}$) nachweisbar als bei Schülern, welche im herkömmlichen Unterricht extrem häufig "unterrichtsbezogene Aktivität" aufweisen²⁾ (vgl. 2.23 und 5.20).

6.50 "Konzentrativer Arbeitshaltung"³⁾

H 19 : Der Zusammenhang von "konzentrativer Arbeitshaltung" und der Lernleistung in den beiden Lernzielkategorien ist bei beiden Unterrichtsmethoden signifikant.

H 20 : Der Zusammenhang von "konzentrativer Arbeitshaltung" und Lernleistung ist im herkömmlichen Unterricht signifikant größer als bei Programmierter Unterweisung.

-
- 1) Darunter fassen wir alle Schüler, welche aufgrund der Daten unserer Beobachtungsprotokolle im unteren Fünftel der Häufigkeitsverteilung liegen.
 - 2) Darunter fassen wir alle Schüler, welche aufgrund der Daten unserer Beobachtungsprotokolle im oberen Fünftel der Häufigkeitsverteilung liegen.
 - 3) Vgl. 7.824

6.60 "Einstellung zur Schule"¹⁾

H 21 : Der Zusammenhang zwischen der "Einstellung zur Schule" und der Lernleistung" ist bei beiden Lernzielkategorien im herkömmlichen Unterricht signifikant größer als bei Programmierter Unterweisung.

6.70 "Einstellung zum Lehrer"²⁾

H 22 : Der Zusammenhang zwischen der "Einstellung zum Lehrer" und der Lernleistung ist bei beiden Lernzielkategorien im herkömmlichen Unterricht signifikant größer als bei Programmierter Unterweisung.

1) Vgl. 7.822

2) Vgl. 7.823

7.00 Methode der Untersuchung

7.10 Grundsätzliche Probleme der experimentellen Forschung im pädagogischen Bereich

Das Experiment ist das einzige Mittel, strittige Fragen der pädagogischen Praxis zu entscheiden, die einzige Möglichkeit, pädagogische Verbesserungen zu rechtfertigen und der einzige Weg, "eine Grundlage für eine anwachsende Tradition zu schaffen, in der Verbesserungen eingeführt werden können, ohne daß die Gefahr besteht, frühere Erkenntnisse zugunsten minderwertiger neuer Ideen willkürlich fallen zu lassen."¹⁾ Diese früher heftig umstrittene Grundhaltung wird auch heute noch keineswegs durchgehend akzeptiert. Dies hat vielfältige Gründe. Wesentlich trug dazu einmal die vielerorts beobachtbare Überschätzung experimentell begründeter Aussagen bei, was häufig gleichzeitig mit einer übertriebenen Mißachtung nichtexperimenteller Arbeiten verbunden ist. Zum anderen zeigte sich im Verlauf der experimentellen Arbeit, daß deren Durchführung oft langwierig und aufwendig war und trotzdem keine eindeutigen Befunde ergab²⁾. Ernüchternd war und ist vor allem die

-
- 1) E. SCHWARZ: "Experimentelle und quasiexperimentelle Anordnungen in der Unterrichtsforschung", in: K.-H. INGENKAMP (Hrsg.): "Handbuch der Unterrichtsforschung Teil I", Weinheim-Berlin-Basel 1970, S. 453
 - 2) Das gilt insbesondere für die im Rahmen der Pädagogik aus sachimmanenten Gründen häufig eingesetzten "feldexperimentellen Studien", welche meistens mit einem großen organisatorischen Aufwand verbunden sind und aufgrund der schwerer als im Labor zu kontrollierenden Störvariablen häufig einen hohen Anteil an Fehlervarianz aufweisen, welcher die untersuchten Effekte verdeckt bzw. vortäuscht.

Vielfalt sich widersprechender Befunde. Häufig zeigen sich eindeutige und wiederholt bestätigte Ergebnisse nur dort, wo bereits vorwissenschaftlich erarbeitete Aussagen vorliegen. Die wesentliche Ursache für die unbefriedigenden Ergebnisse experimenteller Ansätze im pädagogischen Bereich liegt wohl primär darin, daß ausgesprochen komplizierte Fragestellungen mit einem unzulänglichen methodischen Instrumentarium angegangen worden sind. Mangelhafte Versuchsplanung und Verrechnungstechniken, für welche häufig die Voraussetzungen fehlen sowie ungenügendes Wissen um mögliche Fehlerquellen führten zu falschen, zumindest jedoch widersprüchlichen Ergebnissen sowie deren falscher Interpretation.

"Das traditionelle methodische Verfahren besteht darin, zwei oder mehr Gruppen von Schülern so auszuwählen, daß jede Gruppe soweit als möglich repräsentativ für ein zuvor definiertes, einheitliches Gruppenkollektiv von Schülern bzw. Schulen ist. In jeder Gruppe wird sodann eine von den zu vergleichenden Methoden angewandt¹⁾. Voraussetzung für die Herstellung der Repräsentativität und der Gleichheit der in der Untersuchung verwendeten Stichproben ist jedoch eine echte Zufallsauslese. Für Laborexperimente ist dieses Problem ohne größere Schwierigkeiten lösbar, schwierig wird es dagegen bei Untersuchungen im Rahmen der "schulischen Praxis". Hier werden insbesondere zwei Gründe genannt, welche einer echten Zufallsauslese im Wege stehen. Einmal ist es häufig aus organisatorischen Gründen notwendig, auf geschlossene Gruppen in Form von Klassen zurückzugreifen, da einer

1) F. SÜLLWOLD: "Zur Frage der Präzision und Ergiebigkeit von pädagogischen Experimenten", in: "Pädagogische Forschung und Pädagogische Praxis", Veröffentl. d. Hochsch. f. intern. päd. Forschung, 1, 1958, S. 187

Zufallsauslese oft unüberwindliche Schwierigkeiten entgegenstehen¹⁾. Gewichtiger scheint jedoch das zweite Argument zu sein. Es fragt sich nämlich, "ob der soziale und leistungsmäßige Zusammenhang einer Klasse im Interesse einer experimentell geforderten Stichprobenverteilung einfach aufgeteilt werden kann, ohne wiederum eine Verletzung der pädagogischen Situation herbeizuführen ... Klassen, insbesondere dann, wenn sie sich als pädagogische Gruppe organisiert haben, sind eben noch etwas anderes als Experimentiergruppen, nämlich in ihrer Leistungs- und Sozialstruktur äußerst differenzierte und störbare Gebilde. Werden sie experimentell manipuliert, besteht die Gefahr, daß man schiefe Ergebnisse erhält"²⁾. Die ohne Rücksicht auf die "Störbarkeit des pädagogischen Feldes"³⁾ an nach dem Prinzip der Zufallsauslese gebildeten Stichproben gewonnenen Ergebnisse sind insofern "schief", als eine Generalisierung auf die schulische Praxis nicht oder nur sehr bedingt möglich ist. Solche Untersuchungen können zwar intern valide sein, die externe Validität kann dagegen völlig verlorengehen. Wir müssen in der Tat davon ausgehen,

1) Vgl. P. STRITTMATTER: "Präsentationsmodi von Lehrprogrammen", Weinheim 1970, S. 176

2) G. DIETRICH: "Bildungswirkungen des Gruppenunterrichts", München 1969, S. 55. In demselben Zusammenhang schreibt WINNEFELD: "Wir können demnach als erstes Forschungsprinzip das der Wahrung der pädagogischen Eigenart pädagogischer Felder aufstellen ..." (F.W. WINNEFELD: "Pädagogischer Kontakt und Pädagogisches Feld", München 1957, S. 31)

3) F.W. WINNEFELD, a.a.O., S. 30

daß eine Randomisierung bei den meisten pädagogischen Fragestellungen, sofern die Sozialstruktur von Klassen mit einbezogen werden muß, nicht möglich ist. Die überwiegende Anzahl der in diesem Bereich durchgeführten Untersuchungen begnügte sich angesichts dieser Tatsache mit einer "Parallelisierung" der verglichenen Stichproben nach bestimmten Merkmalen, welche die Untersuchungsergebnisse beeinflussen können¹⁾. Solche Merkmale sind z.B. Alter, Geschlecht, Intelligenz, Schulleistung, Vorwissen usw. Die Auswahl der Parallelisierungskriterien erfolgt dabei mehr oder weniger willkürlich, beachtet werden i.allg. diejenigen Variablen, bei denen man aufgrund bisheriger Untersuchungen einen Einfluß auf die jeweiligen abhängigen Variablen festgestellt hat oder ~~einen~~ solchen jedenfalls vermuten kann. Aus ökonomischen Gründen beschränkt man sich dabei meist auf eine oder zwei Variablen²⁾, die anderen werden als Störvariablen mehr oder weniger exakt (häufig sogar gar nicht) beschrieben³⁾, wobei man meist stillschweigend

-
- 1) An dieser Stelle sei noch einmal ausdrücklich darauf hingewiesen, daß es sich dabei i.allg. nicht um eine echte Parallelisierung handelt, wie sie z.B. MITTENECKER schildert (E. MITTENECKER: "Planung und statistische Auswertung von Experimenten", Wien 1963, S. 29 f. und 115 ff.) sondern meist nur darum, daß bezüglich bestimmter Kriteriumsvariablen festgestellt wird, daß sich die arithmetischen Mittel und (bei sorgfältig konzipierten Arbeiten) die Streuungen der verglichenen Gruppen nicht signifikant unterscheiden. Es braucht an dieser Stelle nicht ausgeführt zu werden, daß bei einem solchen Verfahren mit beträchtlicher Wahrscheinlichkeit davon ausgegangen werden kann, daß Versuchs- und Kontrollgruppe nicht derselben Population entstammen.
 - 2) Bei Vergleichsstudien werden hier fast durchgehend "Intelligenz" und "Vorwissen" herangezogen, wobei letzteres häufig über Lernzuwachsformeln kontrolliert wird.
 - 3) Hier werden besonders häufig "Leistungsmotivation", "Anspruchsniveau", "Interesse" und "Lehrstil" erwähnt.

davon ausgeht, daß sie sich in ihrer Wirkung auf die untersuchten Variablen weitgehend aufheben. Bei einer echten Zufallsstichprobe, sofern sie nur groß genug ist, kann man tatsächlich davon ausgehen, daß sich die durch solchermaßen in Kauf genommene Störvariablen verursachte Störvarianz weitgehend aufhebt, bei Klassen und Schulen dagegen existieren häufig systematische, experimentell nicht oder nur unzulänglich faßbare Unterschiede¹⁾. Einen wesentlichen Beitrag in diesem Zusammenhang bietet die Arbeit von ECKEL²⁾. Er geht davon aus, daß sich die "Klassenschülerleistung" aus den Komponenten "Methodeneffekt", "Klasseneinfluß" und einer "klassenunabhängigen Schülerkomponente" zusammensetzt. Dabei bewirken die "Klasseneinflüsse" und die "reine Schülerleistung" Abweichungen von dem interessierenden Methodeneffekt. "Da die Nichtbeachtung des Klasseneffektes scheinbare Methodenunterschiede produziert, braucht man sich nicht darüber zu wundern, daß bei den heute so beliebten Vergleichen zwischen konventionellem und programmiertem Unterricht widersprüchliche Ergebnisse an der Tagesordnung sind ... Die Nichtbeachtung des Klasseneffekts führt in der Regel zu merkwürdigen, bei Methodenvergleichen zu beachtlichen Fehlern"³⁾. Aufgrund einer sorgfältigen Analyse kommt der Autor zu

-
- 1) "Die Tatsache, daß Gruppen geschlossene 'Blöcke' sind und voneinander systematisch abweichen, muß vorher mittels einfacher Varianzanalyse einer größeren Zahl von Gruppen festgestellt werden. Für Schulen ist dies aber schon oft bewiesen worden" (E. MITTEN-ECKER, a.a.O., S. 148).
 - 2) Vgl. K. ECKEL: "Die Bedeutung des Klasseneffekts für die Schulpädagogische Forschung", in: pl, 3, 1969
 - 3) K. ECKEL, a.a.O., S. 101

dem Schluß, daß bei Zugrundelegung von geschlossenen Klassen mindestens 100 Klassen notwendig seien, um sichere statistische induktive Schlüsse von der Stichprobe auf die betrachtete Gesamtheit ziehen zu können¹⁾. Dabei geht ECKEL davon aus, daß die Klassen tatsächlich nach dem Zufall ausgewählt worden sind. Seinen Ausführungen liegt der übliche einfache Vergleich zwischen Versuchs- und Kontrollgruppe zugrunde. Wir werden bei der Besprechung des von uns gewählten Versuches darauf eingehen, ob es keinen ökonomischeren Weg als den von ECKEL aufgezeigten gibt, um zu einer ausreichenden Aussagegenauigkeit zu gelangen²⁾.

Die oben angeführte Praxis, mit Hilfe einiger wesentlicher Kriterien eine "Parallelisierung" durchzuführen,

-
- 1) Zu entsprechenden Zahlen gelangte man auch in den Vereinigten Staaten, wie aus dem Bulletin Nr. 2 der International Association of Educational Achievement, Mai 1967, S. 18, entnommen werden kann (vgl. dazu K. ECKEL, a.a.O., S. 102).
 - 2) ECKEL zieht aus seinen Forderungen auch die entsprechenden Konsequenzen, indem er darauf hinweist, daß ein solcher Arbeitsaufwand nur noch mittels elektronischer Datenverarbeitung zu bewältigen sei. "Ohne Datenverarbeitung ist erfahrungswissenschaftliche Pädagogik nicht möglich" (K. ECKEL: "Von der empirischen zur erfahrungswissenschaftlichen Pädagogik", in: pl, 4, 1970, S. 209). Der Forderung von ECKEL nach ca. 100 Klassen zur experimentellen Klärung schulpädagogischer Fragen wurde andernorts widersprochen. "Die geforderte 'Größenordnung hundert' erscheint nicht nur willkürlich; bei ökonomischer Versuchsplanung sollte man diese Größenordnung ganz erheblich unterschreiten, um nicht Gefahr zu laufen, so minimale 'Effekte' nachzuweisen, daß sie für die Pädagogik praktisch bedeutungslos sind" (K.J. KLAUER: "Die Bedeutung des Klasseneffektes für die Schulpädagogische Forschung", in: pl, 3, 1970). KLAUER geht sogar davon aus, daß es möglich ist, bereits im Rahmen einer Klasse sinnvolle Forschungsarbeiten anzusetzen, sofern es sich um ein theorieorientiertes Problem handelt (vgl. K.J. KLAUER: "Über Möglichkeiten der Experimentalforschung in Schulklassen", in: pl, 1, 1971, S. 2 ff.).

ist ungenügend und dies ist umso bedenklicher, als häufig nicht einmal bei den methodischen Vorerwägungen auf die oben geschilderte Problematik eingegangen wird. Obwohl die oben geschilderte Unzulänglichkeit einer solchermaßen durchgeführten "Parallelisierung" schon früh erkannt worden ist und entsprechend modifizierte Versuchspläne angeboten worden sind, wurde in der einschlägigen pädagogischen Forschung kaum Gebrauch davon gemacht¹⁾. Ein solchermaßen modifizierter Versuchsplan zeichnet sich dadurch aus, daß alle Probanden bzw. alle Gruppen allen Behandlungen ausgesetzt werden. Solche Versuchspläne laufen unter der Bezeichnung "rotation experiments"²⁾. Auf solche Versuchspläne wurde häufig selbst dann verzichtet, wenn keine sachlichen oder organisatorischen Schwierigkeiten zu ihrer Realisation bestanden. Manche Autoren hielten es allerdings immerhin für notwendig, auf die einschlägigen Versuchspläne hinzuweisen, sie verzichteten aber auf deren Anwendung mit dem Argument, daß eine beliebige Rotation nicht möglich sei, ohne die Natürlichkeit der pädagogischen Situation wiederum empfindlich zu stören³⁾. Unter Verweis auf die externe Validität wird hier der Aspekt der internen Vali-

-
- 1) So wies bereits 1923 W.A. Mc. CALL ("How to Experiment in Education", New York 1923) auf die Möglichkeit des Rotationsexperiments im Rahmen der Unterrichtsforschung hin.
 - 2) Häufig werden auch die Ausdrücke "counterbalanced designs", "cross-over designs" und "switch-over designs" verwendet (vgl. E. SCHWARZ, a.a.O., S. 570)
 - 3) Vgl. G. DIETRICH, a.a.O., S. 54

dität z.T. erschreckend vernachlässigt. Es soll hier nicht bestritten werden, daß die Rotation der Versuchsbedingungen bzw. der Versuchsgruppen ein Verfahren darstellt, das bei manchen Fragestellungen kaum oder gar nicht anwendbar ist¹⁾, in jedem Fall aber sorgfältiger methodischer Überlegungen und Vorkehrungen bedarf, um den Vorteil dieser Methode auszunutzen, ohne durch deren Anwendung neue Fehlerquellen zu schaffen, welche ihrerseits die Validität der Ergebnisse beeinträchtigen. Wenn Autoren allerdings aufgrund der Befürchtung, daß eine Rotation der Versuchsbedingungen die "Echtheit der pädagogischen Situation" zerstöre, auf dieselbe verzichten und auch andere vorteilhaftere "experimental designs" bzw. Verrechnungsmethoden nicht verwenden und statt dessen auf die einfache quasi-experimentelle Anordnung von Versuchs- und Kontrollgruppe, wie wir sie oben beschrieben haben, zurückgreifen, dann sollten sie jedenfalls wenigstens davon ausgehen, daß solchermaßen angesetzte Untersuchungen ohne Wiederholung unter gleichen Bedingungen mit denselben Ergebnissen keine Ausgangsposition für eine solide theoretische Diskussion der gewonnenen Daten ergeben. In

1) So ist es durchaus nachvollziehbar, wenn DIETRICH a.a.O., S. 55) schreibt: "... weder vermag eine gruppenunterrichtlich geführte Klasse sich einfach auf andere Unterrichtsformen umzustellen oder kann eine frontalunterrichtlich geführte Klasse nach Belieben sich gruppenunterrichtlich organisieren, noch besteht die Möglichkeit, daß ein Lehrer, der bisher Frontalunterricht pflegte, nun auf einmal in der experimentellen Situation Gruppenunterricht betreibt."

diesem Sinne sind auch die Ausführungen von EIGEN zu verstehen, wenn er schreibt: "In vielerlei Hinsicht ist die Wiederholbarkeit das wichtigste Kriterium für die Beurteilung von wissenschaftlichen Untersuchungen. Doch gibt es auf dem Gebiet des Programmierten Unterrichts keine einzige Untersuchung, von der der Verfasser dieses Artikels weiß, daß sie wiederholt worden ist. ... Bevor im Bereich des Programmierten Unterrichts keine Wiederholung von Experimenten stattfindet, haben wir nur geringe Verlässlichkeit, und demzufolge werden wir bei zukünftigen Untersuchungen behindert sein, da wir durch nicht wiederholbare Experimente irreführt werden können"¹⁾. Fünf Jahre nach der Veröffentlichung von EIGEN hat sich das Bild noch nicht wesentlich geändert. Es wurden zwar einige partielle Wiederholungsuntersuchungen durchgeführt²⁾, diese sind jedoch noch äußerst selten. Zum gegenwertigen Zeitpunkt kann mit Sicherheit konstatiert werden, daß eine Vielzahl an widersprüchlichen Ergebnissen auf die Stichprobenverzerrungen, welche aufgrund der unzulänglichen "Parallelisierung" entstanden, zurückgeführt werden müssen. Solche Stichprobenverzerrungen wirken sich dabei sowohl auf die interne als auch die externe Validität aus. Im letzteren Fall treten Wechselwirkungen zwischen der verzerrten Stichprobenauswahl und den experimentellen Variablen auf³⁾. Wenn an dieser Stelle die Problematik der Stichprobenverzerrung besonders betont wird, soll darüber nicht vergessen werden, daß auch andere Störvariablen bei pädagogischen Experimenten einen hohen Varianzanteil aufweisen können. Wir werden im folgenden darauf ausführlich zu sprechen kommen.

-
- 1) L.D. EIGEN: "Probleme der Forschung im Bereich des Programmierten Unterrichts", in: L.J. ISSING (Hrsg.): "Die Programmierte Unterweisung in den USA heute", Weinheim 1967, S. 83 f.
 - 2) Vgl. H. WALTER: "Das Problem der Validierung von Unterrichtsprogrammen", in: G. DIETRICH: "Beiträge zur Unterrichtspsychologie der Hauptschule", erscheint 1971
 - 3) Vgl. E. SCHWARZ, a.a.O., S. 461

7.20 Störvariablen, welche die interne und externe Validität beeinträchtigen

Im folgenden sollen wesentliche Faktoren, welche die innere und äußere Validität experimenteller und quasi-experimenteller Anordnungen¹⁾ bestimmen, diskutiert werden. Die Diskussion der internen Validität gilt dabei der Frage, inwieweit die vom Versuchsleiter gesetzten experimentellen Bedingungen (unterschiedliche Behandlungen) wirklich unterschiedliche Effekte verursachen bzw. inwieweit Störvariablen, welche im experimentellen Aufbau selbst begründet liegen, unterschiedliche Effekte verursachen können, die nicht von den unabhängigen Variablen bewirkt sind. Die Diskussion der äußeren Gültigkeit gilt der Frage nach der Generalisierbarkeit von Untersuchungsergebnissen. Wenngleich die innere Gültigkeit Voraussetzung für eine Generalisierbarkeit überhaupt ist, können durch den experimentellen Aufbau Situationen geschaffen werden, welche zwar die innere Validität nicht beeinträchtigen, trotzdem aber eine Generalisierung stark beeinträchtigen. Da die interne Validität unabdingbare Voraussetzung für eine Generalisierung überhaupt ist, muß ihr jedoch der Vorrang eingeräumt werden, ohne daß dabei die Forderung nach einer optimalen Konstellation für die externe Validität vernachlässigt werden sollte. An den im folgenden vorgestellten Kriterien soll später die Validität unseres eigenen Versuchsplans gemessen werden. Wir halten uns daher bei der Erörterung an diejenigen Bedingungen, welche für eine Vergleichsuntersuchung, wie sie auch von

1) Darunter werden alle Versuchspläne gefaßt, welche mit nicht-randomisierten Stichproben arbeiten (vgl. E. SCHWARZ, a.a.O., S. 452).

uns durchgeführt worden ist, kennzeichnend sind, d.h. daß wir nur diejenigen Störvariablen aufführen werden, welche für unser Anliegen von wesentlicher Bedeutung sind.¹⁾

7.21 Faktoren, welche die innere Gültigkeit beeinträchtigen

a) Zwischenzeitliches Geschehen

Darunter werden alle äußeren Ereignisse verstanden, welche zusätzlich zu den experimentell gesetzten unabhängigen Variablen während dem zeitlichen Verlauf einer Untersuchung im Hinblick auf die abhängigen Variablen wirksam werden. Bei der Diskussion der Effektivität verschiedener Unterrichtsmethoden wird häufig die Meinung vertreten, daß die Wirksamkeit einer neuen Unterrichtsmethode erst dann wirklich zuverlässig beurteilt werden kann, wenn die Schüler über einen längeren Zeitraum hinweg der neuen Methode ausgesetzt seien, da nur so der i.allg. leistungsfördernde "Neuigkeitseffekt" ausgeschaltet werden könne. Weiter wird ins Feld geführt, daß, um mit einer neuen Unterrichtsmethode optimal zu arbeiten, eine relativ lange Eingewöhnungsphase notwendig sei. Auf diesem Hintergrund sind die Längsschnittstudien zu sehen, welche in letzter Zeit auf dem Gebiet der Programmierten Unterweisung auch in Deutschland durchgeführt worden sind. Gegner der Programmierten Unterweisung führen häufig ins Feld, daß die negativen Auswirkungen dieser

1) Eine erschöpfende Zusammenstellung derselben findet sich bei E. SCHWARZ, a.a.O., S. 460 ff.

neuen Unterrichtsmethode erst über einen längeren Zeitraum hinweg sichtbar würden. Ungeachtet der Tatsache, daß dem "Novitätseffekt" und der Notwendigkeit einer längeren Gewöhnungsphase (jedenfalls teilweise) vorexperimentell Rechnung getragen werden kann, muß in unserem Zusammenhang darauf hingewiesen werden, daß die Zahl der zwischenzeitlich bedingten Störvariablen umso größer wird, je länger der Zeitraum ist, über welchen hinweg ein Experiment sich hinzieht. Die Gefahr außerexperimenteller Einflüsse ist gerade bei schulischen Unterrichtsstoffen außerordentlich groß, sei es durch die bislang viel zu wenig beachtete Auswirkung auf die abhängige Variable "Lernleistung" durch Unterrichtsstoffe, welche in verwandten Unterrichtsfächern gleichzeitig behandelt werden, sei es durch außerschulische Einflüsse wie Fernsehen, Tageszeitung etc. Wenn an dieser Stelle besonderer Wert auf außerexperimentelle Informationseinflüsse gelegt wird, soll darüber nicht vergessen werden, daß es noch eine Vielzahl anderer zwischenzeitlich bedingter Störvariablen gibt. Außerexperimentelle Informationseinflüsse stellen allerdings bei Experimenten, welche als abhängige Variable eine Schülerleistung in irgend einem Schulfach messen, einen besonders bedeutsamen Faktor dar¹⁾. Wir sind daher der Ansicht, daß im Hinblick auf die interne Validität Vergleichsuntersuchungen nicht in Form von Längsschnittstudien angelegt werden sollten. Dies deckt sich übrigens auch mit der Forderung nach ökonomischen Versuchsplänen. Wenn aufgrund spezieller Fragestellungen Experimente angesetzt werden müssen, welche über ein Jahr hin und länger laufen, darf nicht vergessen werden, daß die

1) Dies gilt insbesondere dann, wenn sich eine Untersuchung auf die Leistungen im Fach Geschichte bzw. Sozialkunde bezieht.

zwischenzeitlich bedingten Störvariablen wenigstens durch eine sehr große Stichprobe "ausgeglichen" werden sollten. Bei einer sehr großen, zufällig ausgewählten, Stichprobe kann nämlich mit hoher Wahrscheinlichkeit davon ausgegangen werden, daß zwischenzeitliche Ereignisse lediglich eine unsystematische Störvarianz bewirken. Darüber hinaus sollten experimentell nicht behandelte Kontrollgruppen unbedingt in den Versuchsplan mit aufgenommen werden.¹⁾

b) Testen

Darunter wird die Auswirkung einer Testdurchführung auf eine weitere Testdurchführung verstanden. Schüler, welche einen Leistungs- bzw. Intelligenztest wiederholt bearbeiten, schneiden i.allg. besser ab, als solche, welche einen Test nur einmal bearbeiten²⁾. Da bei Vergleichsstudien häufig mit sog. Lernzuwachsformeln, welche alle auf einem identischen Vor- und Schlußtest basieren, gearbeitet wird³⁾, sollte die Reaktivität der Meßbedingungen im Auge behalten werden. Allerdings

-
- 1) Die in diesem Kapitel gemachten Ausführungen gelten weithin auch für die Störvariable "Reifung", das sind "biologische und psychische Vorgänge, ... die sich systematisch und unabhängig von besonderen äußeren Ereignissen im Laufe der Zeit ändern" (E. SCHWARZ, a.a.O., S. 466) und dadurch die Versuchsergebnisse verzerren.
 - 2) Bei Schülern mit geringer Testerfahrung macht diese Verbesserung bei Intelligenztests im Mittel 3 bis 5 IQ-Punkte aus (vgl. E. SCHWARZ, a.a.O., S. 466 ff.).
 - 3) Vgl. K. HEIPKE: "Die Bedeutung mathematischer Lernmodelle bei der Auswertung von Lernversuchen", in: pl, 3, 1968, S. 98

spielt die Reaktivität von Meßinstrumenten bei Leistungsexperimenten i.allg. eine weit geringere Rolle als dies z.B. bei sozialpsychologischen Untersuchungen der Fall ist, welche mit Einstellungsmaßen arbeiten.¹⁾ Zudem muß bei dieser Störvariable vermerkt werden, daß sie zwar grundsätzlich Methodeneffekte vortäuschen kann, dies gilt jedoch für Versuchs- und Kontrollgruppe i.allg. gleichermaßen, so daß der Vergleich der beiden Gruppen nicht unmittelbar tangiert wird. Voraussetzung dafür sind allerdings wirklich vergleichbare Stichproben, eine Bedingung, welche gerade im Rahmen von Vergleichsuntersuchungen in den seltensten Fällen erfüllt ist.

c) Hilfsmittel

Hierunter werden Änderungen in der Beschaffenheit der Meßinstrumente oder Änderungen der Beobachter und Auswerter verstanden, welche Änderungen der erhaltenen Messungen hervorrufen. Bei Vergleichsuntersuchungen kann man i.allg. davon ausgehen, daß die Leistungsprüfungen objektiv auswertbar sind. Sofern im Rahmen solcher Untersuchungen als abhängige Variable Beobachtungsdaten verwendet werden, kann hier - insbesondere wenn diese nur ungenügend operationalisiert sind - eine wesentliche Quelle für systematische Fehlervarianz liegen.

d) Statistische Regression

Hierunter werden statistische Artefakte verstanden, welche experimentelle Einflüsse vortäuschen. Dieses Problem

1) Vgl. C. WINDLE: "Test-retest effect on personality questionnaires", in: Educ. psychol. Measmt., 1954, 14, S. 617-633

stellt sich vor allem bei der Verwendung von sog. Extremgruppen. Es wurde bei einer Vielzahl von Untersuchungen nicht beachtet und führte dementsprechend zu falschen Schlüssen. Wenn z.B. die Wirksamkeit einer Unterrichtsmethode an einer Schülergruppe demonstriert werden soll, welche in einem experimentbezogenen Schulleistungstest besonders schlecht abgeschnitten hat, wird das arithmetische Mittel dieser Gruppe bei einer zweiten Messung auch ohne experimentelle Beeinflussung i.allg. höher liegen, da die Wahrscheinlichkeit, daß die Schüler wiederum eine so schlechte Leistung wie bei der ersten Messung aufweisen, relativ gering ist. Je schlechter die ursprüngliche durchschnittliche Schülerleistung ist, desto geringer ist die Wahrscheinlichkeit, daß sie bei einer zweiten Messung ohne dazwischenliegende experimentelle Behandlung wiederum so gering ist, d.h. aber umgekehrt, desto größer ist die Wahrscheinlichkeit, daß bei einer wiederholten Messung ein besseres Ergebnis erzielt wird.¹⁾ Da nun aber zwischen der ersten und zweiten Messung eine experimentell gesetzte Beeinflussung liegt, neigt man leicht dazu, die stattgefundenen Veränderungen auf die äußeren experimentellen Einflüsse zurückzuführen²⁾. Um

1) Vgl. E. SCHWARZ, a.a.O., S. 471 ff.

2) Entsprechende Trugschlüsse finden sich in großer Zahl. Hierher gehört z.B. das häufig diskutierte Problem, daß die intelligentesten Schüler in einem Schulleistungstest nicht die besten Leistungen zeigen und umgekehrt. Daraus wird nun gefolgert, daß die Schule den (intelligenten) Schülern nicht genügend Anregung gäbe, oder daß die Höhe einer Leistung nicht ausschließlich von der Intelligenzausprägung abhängt. Unbesehen der möglichen psychologischen Relevanz solcher Annahmen handelt es sich hier zunächst einmal schlicht um eine Auswirkung der statistischen Regression, zumindest können divergierende Befunde, wie sie oben geschildert wurden, nicht ohne Berücksichtigung des Phänomens der statistischen Regression interpretiert werden.

den Einfluß der Regression bei der Verwendung von Extremgruppen grob abschätzen zu können, muß die Korrelationshöhe der beiden verglichenen Meßwertreihen beachtet werden. Je größer der Korrelationskoeffizient ist, desto geringer ist die Regression auf den Mittelwert. Regressionseffekte sind bei Extremgruppen eine unvermeidbare Erscheinung, da sie in jedem Fall allein durch eine unvollkommene Testwiederholungskorrelation verursacht werden. Insbesondere bei den im Rahmen von Vergleichsstudien häufig herangezogenen Extremgruppen aus dem oberen und unteren Intelligenzbereich muß die Regression zum Mittelwert sorgfältig analysiert werden, bevor eine Interpretation vorliegender Untersuchungsergebnisse ansetzen kann.¹⁾

e) Stichprobenverzerrungen

Hierunter wird derjenige Anteil an Störvarianz verstanden, welcher durch Nichtübereinstimmung der verglichenen Stichproben verursacht wird. Die am weitesten verbreitete quasiexperimentelle Anordnung enthält - wie wir oben mehrfach gesehen haben - eine Experimental- und eine Kontrollgruppe, wobei beide nach bestimmten

1) So müßten z.B. alle Untersuchungen über die Wirksamkeit des Programmierten Lernens bei Sonderschülern bzw. im weiteren Sinn bei minderbegabten Kindern überhaupt auf den Einfluß der statistischen Regression hin untersucht werden. Gerade von dieser Seite hört man häufig das Argument, daß unintelligente Extremgruppen bei Programmierter Unterweisung besonders gute Lernfortschritte aufweisen. Hier liegt zunächst der Verdacht nahe, daß es sich bei diesen Untersuchungsergebnissen - jedenfalls teilweise - um statistische Artefakte handelt.

Kriterien "parallelisiert" sind. Deren Vergleichbarkeit ist jedoch aufgrund der Tatsache, daß es sich nicht um Zufallsstichproben handelt und darüber hinaus nur eine Scheinparallelisierung vorgenommen wird, nicht nur nicht gewährleistet, es kann darüber hinaus sogar mit an Sicherheit grenzender Wahrscheinlichkeit davon ausgegangen werden, daß die beiden Stichproben voneinander systematisch abweichen. Selbst im Falle einer echten Parallelisierung können die Stichproben nicht im Sinne einer Randomisierung aneinander angeglichen werden¹⁾. Wenn sich in einer solchen Anordnung "die Mittelwerte der Gruppen wesentlich unterscheiden, kann durch Parallelisierung die beabsichtigte Gleichheit nicht hergestellt werden; darüber hinaus treten mit Sicherheit unerwünschte Regressionseffekte auf. Die beiden Gruppen werden sich deshalb mit Sicherheit in ihren Nachtestergebnissen völlig unabhängig von einem Einfluß von X (gemeint ist die Einwirkung der experimentell gesetzten unabhängigen Variable (der Verf.)) unterscheiden. Der Unterschied wird proportional zu dem Unterschied zwischen den Populationen, aus denen die Auswahl erfolgte, variieren und umgekehrt zur Testwiederholungskorrelation"²⁾. Zusätzlich zu den oben geschilderten Auswirkungen einer verzerrten Stichprobenauswahl muß noch erwähnt werden, daß relativ häufig eine Wechselbeziehung zwischen Reifung und Stichprobenauswahl beobachtet wurde, welche Veränderungen bewirkte, die auch ohne experimentellen Einfluß aufgetreten waren, irrtüm-

1) Vgl. E. SCHWARZ, a.a.O., S. 566

2) Vgl. E. SCHWARZ, a.a.O., S. 567

licherweise aber für einen solchen gehalten wurden.¹⁾
Grundsätzlich sind auch Wechselwirkungen anderer Art, z.B. zwischen Stichprobenauswahl und Testen oder Stichprobenauswahl und Hilfsmitteln denkbar. Solche Stichprobenverzerrungen stellen im Rahmen von Vergleichsuntersuchungen wohl die entscheidende Störvariable dar, da - wie wir oben gesehen haben - gezwungenermaßen meistens auf natürliche Klasseneinheiten zurückgegriffen werden muß. Die bislang behandelten Störvariablen "zwischenzeitliches Geschehen", "Testung" und "Hilfsmittel" können alle durch Einbezug einer Kontrollgruppe relativ sauber kontrolliert werden. Die statistische Regression kann durch Vermeidung von Extremgruppen wenigstens teilweise umgangen werden. Demgegenüber scheint eine Stichprobenverzerrung im Rahmen von Vergleichsuntersuchungen aufgrund der Verwendung natürlicher Klasseneinheiten unvermeidbar zu sein, es sei denn, man verzichtet von vornherein weitgehend auf die externe Validität. Noch gewichtiger wird das Problem der Stichprobenverzerrung, wenn man an die Vielzahl von Wechselwirkungsmöglichkeiten denkt und darüber hinaus berücksichtigt, daß die in die Untersuchung einbezogenen Klassen nicht nur aufgrund der verschiedenen Schülerpersönlichkeiten divergieren, sondern darüber hinaus durch Eigenschaften, welche den Klassen jeweils in ihrer Eigenschaft als individueller sozialer Verband zukommen. Die Stichprobenverzerrung stellt also im Rahmen von Vergleichsstudien eine Störvariable dar, die automatisch in jede Untersuchung mit eingeht und einen hohen Anteil an systematischer Fehlervarianz bewirkt, es sei denn, man ent-

1) Vgl. E. SCHWARZ, a.a.O., S. 564 f.

scheidet sich für eine echte Randomisierung, welcher - wie wir oben gesehen haben - gewichtige Gründe entgegenstehen. Angesichts dieses Sachverhalts muß gerade in dieser Hinsicht bei der Durchführung einschlägiger Untersuchungen darauf geachtet werden, daß eine experimentelle Kontrolle in den Versuchsplan mit eingebaut wird. Es ist überraschend, daß im Rahmen der bislang durchgeführten Vergleichsuntersuchungen entsprechende Bemühungen kaum sichtbar wurden. Einmal bietet sich zur Kontrolle dieser Störvariable die sog. ausbalancierte Versuchsanordnung an, welche wir oben bereits angesprochen haben¹⁾. Eine weitere Kontrollmöglichkeit bieten die kovarianzanalytischen Verrechnungsverfahren sowie der Rückgriff auf das Gesetz der großen Zahl²⁾. Die Anwendung der Kovarianzanalyse enthält jedoch Voraussetzungen (z.B. Homogenität der Regression), welche bei dem häufig praktizierten Vorgehen mit Experimental- und Kontrollgruppe ohne echte Randomisierung nicht von vornherein gegeben sind³⁾. Es kann nicht bezweifelt werden, daß zukünftige Vergleichsstudien und im weiteren die Unterrichtsforschung überhaupt, auf diese Kontrollmöglichkeiten zurückgreifen müssen⁴⁾.

1) Sie wurde eigens für den Fall entworfen, daß keine Rückgriffsmöglichkeit auf randomisierte Stichproben besteht (vgl. E. SCHWARZ, a.a.O., S. 452)

2) Vgl. K. ECKEL, a.a.O., S. 102

3) Nähere Ausführungen finden sich bei E.F. LINDQUIST: "Designs and analysis of experiments in psychology and education", Boston 1953.

4) Erst in jüngster Zeit erschienen im deutschsprachigen Bereich Untersuchungen, welche diese Bedingung erfüllen (vgl. M. HIRZEL: "Partnerarbeit im Programmierten Unterricht", Stuttgart 1970 und K.J. KLAUER: "Orthographisches Lernen als Funktion der Lehrmethode des Leistungs- und des Intelligenzniveaus", in: ZeF, 2, 1968, 1).

f) Experimentelle Einbußen

Darunter verstehen wir Stichprobenverzerrungen, welche dadurch entstehen, daß im Verlauf des Experiments Probanden in unterschiedlicher Weise die Versuchsgruppen verlassen. Grundsätzlich treten hier dieselben Schwierigkeiten auf, welche wir unter Punkt e) dieses Kapitels geschildert haben, nur daß diese eben nicht bereits durch die Stichprobenauswahl verursacht wurden sondern sich im Verlauf des Experiments einstellen. Solche experimentellen Einbußen treten im Rahmen von Vergleichsuntersuchungen häufig auf. Selbst wenn solche Studien relativ kurzfristig angelegt sind, laufen sie i.allg. wenigstens über einige Tage. Wenn ein Proband auch nur *e i n e* Stunde fehlt, kann er in die (verrechnete) Endstichprobe grundsätzlich nicht mehr aufgenommen werden. Bei solchen feldexperimentell angelegten Untersuchungen muß daher bei der Bestimmung der Größe der Ausgangsstichprobe von Anfang an mit einer relativ hohen Ausfallsquote gerechnet werden, es handelt sich also auch hier um eine Störvariable, welche letzthin unvermeidbar ist. Bei Vergleichsstudien, welche von Anfang an auf randomisierte Stichproben verzichten, spielt die solchermaßen entstandene Stichprobenverzerrung allerdings keine so große Rolle wie bei Verwendung randomisierter Stichproben, da aufgrund der natürlichen Klasseneinheiten die experimentell von Anfang an gesetzte Stichprobenverzerrung so groß ist, daß diese auch durch experimentelle Einbußen nur noch geringfügig beeinflusst werden dürfte. Als Kontrollmöglichkeit bietet sich wiederum das Rotationsexperiment an.

7.22 Faktoren, welche die äußere Gültigkeit beeinträchtigen

a) Wechselwirkung zwischen verzerrter Stichprobenauswahl und den experimentellen Variablen

Darunter werden Einflüsse auf die abhängige Variable verstanden, welche aufgrund der verzerrten Stichprobenauswahl nur für die eingeschränkte Population gelten, aus welcher die untersuchten Stichproben entnommen wurden. Im Rahmen von Vergleichsstudien tritt dieses Problem häufig auf. Wenn z.B. zur Erprobung einer neuen Unterrichtsmethode eine oder mehrere Klassen gesucht werden, so bekommt man häufig von einer Vielzahl von Schulen eine Absage. Diejenigen Schulen, welche sich bereiterklären, sich dem Experiment zu unterziehen, unterscheiden sich möglicherweise von den anderen Schulen in typischer Weise. So kann z.B. angenommen werden, daß diese Schulen in stärkerem Maße als üblich neuen Unterrichtsmethoden gegenüber aufgeschlossen sind. Die solchermaßen gewonnenen Ergebnisse können, wenn überhaupt, auf die Schulen derselben Gattung nicht ohne weiteres übertragen werden, da die gefundenen Effekte möglicherweise gerade auf dem deutlich gesteigerten Interesse und der stärkeren Motivierung der untersuchten Schüler beruhen¹⁾. Sofern man bei solchen Experi-

1) In diesem Zusammenhang schreibt ECKEL: "Mit Hilfe des Prinzips der freiwilligen Beteiligung an Stichprobenuntersuchungen lassen sich in der Regel nur die 'berühmten allerersten Vermutungen' - von denen es in den Gesellschaftswissenschaften nur so wimmelt - gewinnen." (K. ECKEL: "Von der empirischen zur erfahrungswissenschaftlichen Pädagogik", in: pl, 4, 1970, S. 209). In konsequenter Fortführung seiner bereits weiter oben berichteten Vorstellungen fordert er eine "fortwährende Realisierung von Zufallsstichproben" (K. ECKEL, a.a.O., S. 209). Daß diese nicht auf freiwilliger Basis erfolgen kann, ist selbstverständlich.

menten auf freiwillige Mitarbeit von Schulen bzw. Klassen angewiesen ist, sind genaue Angaben über die Zahl und die Art der Schüler unbedingt erforderlich. So kann bei der Interpretation der Versuchsergebnisse wenigstens abgeschätzt werden, ob und inwieweit die aufgrund von Stichprobenverzerrungen entstandenen Wechselwirkungen die Effekte mit verursacht haben könnten¹⁾. Grundsätzlich kann davon ausgegangen werden, daß mit wachsender Ablehnungsquote der zu erwartende Wechselwirkungseffekt immer größer wird.

b) Reaktive Effekte der experimentellen Situation

Hierunter werden jene experimentellen Auswirkungen verstanden, welche eine Generalisierung hinsichtlich solcher Personen unmöglich macht, welche anderen Bedingungen als den in der jeweiligen experimentellen Situation vorherrschenden ausgesetzt sind. Jedes Experiment stellt eine mehr oder weniger künstliche Situation dar. Diese Künstlichkeit und das Wissen der Versuchspersonen, daß sie sich in einer experimentellen Situation befinden, sind im Rahmen von Vergleichsstudien die häufigste Ursache für mangelnde Repräsentativität. Die Wiederholung

1) An dieser Stelle muß ausdrücklich vermerkt werden, daß wir, in Übereinstimmung mit KLAUER (vgl. K.J. KLAUER: "Die Bedeutung des Klasseneffekts für die Schulpädagogische Forschung", in: pl, 3, 1970, S. 149-156) davon ausgehen, daß es durchaus möglich ist, an einigen Klassen (auch unter der Bedingung freiwilliger Mitarbeit) qualifizierte Untersuchungen durchzuführen. ECKELs Position betrachten wir als überzogen (zur Kritik vgl. K.J. KLAUER, a.a.O.).

von lediglich intern validen Versuchen kann das Bemühen, Versuchsbedingungen von möglichst geringer Reaktivität zu schaffen, in keiner Weise ersetzen. Folgende grundsatzlichen Aspekte der Verursachung reaktiver Effekte scheinen uns bei Vergleichsuntersuchungen besonders bedeutsam zu sein:

1. Das Wissen der Schüler, daß sie sich in einer experimentellen Situation befinden.
2. Der äußere organisatorische und technische Aufwand, welchen das Experiment erfordert.
3. Die inhaltliche Gestaltung des Testvorgangs.

zu 1. Nur in ganz wenigen Fällen ist es gelungen, das Wissen des Schülers um das Vorhandensein einer experimentellen Situation zu verhindern¹⁾. Es ist jedoch häufig möglich, die Zielsetzung eines Experiments so zu gestalten, daß eine willentliche Beeinflussung der abhängigen Variable seitens der Versuchspersonen nicht möglich ist. Dies kann einmal einfach dadurch geschehen, daß die Versuchspersonen bezüglich der Zielsetzung des Experiments nicht informiert werden, zum anderen kann darüber hinaus eine irreführende Information gesetzt werden. Trotzdem wird es den Schülern häufig möglich sein, aus der experimentell gesetzten Situation als solcher deren Zielsetzung zu raten.

1) Vorbildlich ist diesbezüglich das experimentelle Vorgehen von PAGE, dem es gelang, innerhalb einer Klasse mit randomisierten Teilstichproben zu arbeiten, ohne daß die Versuchspersonen bemerkten, daß sie sich in einer experimentellen Situation befanden. (Vgl. E.B. PAGE: "Teacher comments and student performance: A seventyfour classroom experiment in school motivation", in: Journ. educ. Psychol., 1958, 49, S. 173-181).

zu 2. Soweit als möglich sollte der apparative und organisatorische Aufwand auf ein Minimum reduziert werden. Bei Schulversuchen sollte z.B. darauf geachtet werden, daß die natürliche räumliche Umgebung erhalten bleibt und die zeitliche Anordnung der experimentellen Beeinflussung sich in den Stundenplan unauffällig einbauen läßt. Eine Aufteilung der Klasse bewirkt mit Sicherheit massive reaktive Effekte.

zu 3. Es liegen genügend Hinweise für die Vermutung vor, daß die inhaltliche Gestaltung des Testvorgangs von entscheidender Bedeutung für die Reaktivität einer experimentellen Situation ist¹⁾. "Mit regulären Klassenarbeiten oder mit Tests, die als reguläre Prüfungen dargeboten werden und im Inhalt den üblichen Arbeiten ähneln"²⁾, kann diese Ursache für reaktive Versuchsbedingungen wahrscheinlich in den meisten Fällen vermieden werden.

7.23 Zusammenfassende Schlußfolgerungen für Methodenvergleiche im Rahmen der Schule

Wir sind schon oben darauf eingegangen, daß die Reduzierung bzw. Eliminierung einer Störvariable nicht automatisch mit einer gleichzeitigen Erhöhung der inneren und äußeren Validität verbunden ist, da durch die Einschränkung einer Störvariable häufig andere Störvariablen begünstigt werden. Dem Entwurf eines Versuchsplans wird daher immer eine sorgfältige Abwägung der Bedingungen für die äußere und innere Validität vorausgehen müssen. Es werden dabei bevorzugt diejenigen Störvariablen zu eliminieren sein, welche die abhängige

1) Vgl. E. SCHWARZ, a.a.O., S. 499 ff.

2) E. SCHWARZ, a.a.O., S. 502

Variable aufgrund theoretischer Erwägungen und empirischer Erfahrung am größten verfälschen. Bei Vergleichsstudien sind dies, wie wir unter 7.10 und 7.20 gesehen haben, in erster Linie die Beeinträchtigung der inneren Validität durch die Stichprobenverzerrung (wobei darunter sowohl die Verzerrung aufgrund der Stichprobenauswahl als auch durch experimentelle Einbußen verstanden wird) und die Reduzierung der äußeren Validität durch reaktive Versuchsbedingungen. Im einzelnen können folgende Punkte hervorgehoben werden, die im Hinblick auf unsere methodischen Erwägungen bevorzugt beachtet werden müssen:

- a) Die Untersuchung soll nicht über einen zu langen Zeitraum laufen (vgl. 7.21 a)).
- b) Bei wiederholten Messungen soll die Reaktivität des Testinstruments beachtet werden (vgl. 7.21 b)).
- c) Die Auswertung der Tests zur Erfassung der abhängigen Variable "Lernleistung" soll möglichst objektiv sein. Verwendete Beobachtungskategorien sollen so formuliert sein, daß eine Änderung der Beobachterurteile bei gleichem Sachverhalt möglichst vermieden wird (vgl. 7.21 c)).
- d) Bei Verwendung von Extremgruppen muß auf das Phänomen der statistischen Regression geachtet werden (vgl. 7.21 d)).
- e) Es müssen Vorkehrungen getroffen werden, um die durch die Stichprobenauswahl und experimentelle Einbußen entstehende Stichprobenverzerrung soweit als möglich zu kontrollieren (vgl. 7.21 e) und 7.21 f)).
- f) Die zur Untersuchung herangezogenen Schulen sollen sich bezüglich ihrer Bereitwilligkeit zur Mitarbeit möglichst nicht von anderen Schulen desselben Schultyps unterscheiden (vgl. 7.22 a)).
- g) Die Schüler sollten die Versuchsergebnisse nicht durch Kenntnis des Versuchsziels bewußt verfälschen können (vgl. 7.22 b)).
- h) Der organisatorische und technische Aufwand soll möglichst unauffällig bleiben (vgl. 7.22 b)).
- i) Das zur Erfassung der abhängigen Variable "Lernleistung" verwendete Instrumentarium soll möglichst demjenigen regulärer Prüfungen gleichen (vgl. 7.22 a)).

7.30 Versuchsplan

7.31 Allgemeine Erörterung verschiedener Versuchspläne

Wir haben bereits an anderer Stelle dieser Arbeit darauf hingewiesen, daß die im Rahmen von Vergleichsstudien übliche quasiexperimentelle Anordnung, d.h. der Rückgriff auf natürliche Klasseneinheiten, von denen ein Teil programmiert, der andere konventionell unterrichtet wird, beträchtliche Fehlerquellen aufweist und zu Ergebnissen führt, die ohne Wiederholungsuntersuchungen kaum zwingend interpretierbar sind. Weiter haben wir darauf hingewiesen, daß die im Rahmen dieser Arbeiten vereinzelt vorgenommene "unechte Parallelisierung", d.h. die Sicherung nichtsignifikanter Unterschiede bezüglich Kriteriumsvariablen wie "Intelligenz" und "Vorwissen", auch nicht annähernd genügen kann, um vergleichbare Stichproben herzustellen. Auch die Versuche einer echten Parallelisierung bei Beibehaltung natürlicher Klasseneinheiten¹⁾ können eine Randomisierung nicht ersetzen²⁾. Da der Rückgriff auf natürliche Klasseneinheiten bei Vergleichsuntersuchungen, welche im

1) Vgl. dazu P. STRITTMATTER: "Präsentationsmodi von Lehrprogrammen", Weinheim 1970, S. 175 ff. Der Autor kennzeichnete jede Versuchsperson durch zwei Vortestmerkmale. "Anhand dieser Merkmale wurden Tripel gebildet, deren Mitglieder jedoch - darin dürfte die entscheidende Schwäche des Verfahrens liegen - nicht per Zufall den 3 Treatment-Gruppen zugeordnet wurden" (P. STRITTMATTER, a.a.O., S. 178).

2) Vgl. E. SCHWARZ, S. 451 f. und 566 f.

Rahmen der Schule durchgeführt werden müssen, nicht nur aus organisatorischen Gründen i.allg. unumgänglich ist, sondern darüber hinaus im Falle einer Zufallsauslese für den herkömmlichen Unterricht Bedingungen geschaffen werden, welche aufgrund ihrer "Künstlichkeit" keinen herkömmlichen Unterricht mehr im Sinne der empirischen Realität, d.h. der pädagogischen Praxis mehr ermöglichen, muß versucht werden, unter Wahrung natürlicher Klasseneinheiten auf andere Versuchsanordnungen zurückzugreifen. Der Vorschlag von ECKEL, auf eine sehr große Zahl (wenigstens 100) an Klassen zurückzugreifen und aus jeder Klasse per Zufall eine Versuchsperson für die Endstichprobe auszuwählen¹⁾, ist nach unserer Auffassung zu aufwendig und würde die Möglichkeiten der Unterrichtsforschung unnötig stark einschränken. Wir haben uns daher angesichts der unter Punkt 7.10 und 7.20 geschilderten zentralen Bedeutung der Stichprobenverzerrung im Rahmen der bisherigen Vergleichsuntersuchungen entschlossen, auf eine ausbalancierte Anordnung zurückzugreifen, bei der die experimentelle Kontrolle dadurch erreicht oder wenigstens die Genauigkeit verbessert wird, daß alle Versuchspersonen allen Behandlungen ausgesetzt werden. Obwohl Mc. CALL das Rotationsexperiment bereits 1923 vorgeschlagen hat²⁾ um eine gewisse Kontrolle zu gewährleisten, wenn eine Zufallsauslese zur Herstellung gleichartiger Stichproben nicht möglich ist, wurde dieser Versuchsplan bzw. entsprechende Variationen in der Unterrichtsforschung bislang kaum verwendet, im Rahmen von Vergleichsuntersuchungen ist uns überhaupt keine ent-

1) Vgl. K. ECKEL: "Die Bedeutung des Klasseneffekts für die Schulpädagogische Forschung", in: pl, 3, 1969, S. 102

2) Vgl. W.A. Mc. CALL: "How to Experiment in Education", New York 1923

sprechende Arbeit bekannt. Wenngleich auch das Rotationsexperiment, solange es auf natürliche Klasseneinheiten zurückgreifen muß, keinen vollwertigen Ersatz für eine Zufallsauslese darstellt, handelt es sich dennoch um eine Versuchsanordnung, welcher nach unserem Dafürhalten im Rahmen von Vergleichsstudien der Vorzug vor der bisher geübten experimentellen bzw. quasi-experimentellen Praxis gebührt.

7.32 Versuchsplan

Unserer Untersuchung liegt folgendes Schema zugrunde

	Programmierter Unterricht	Konventioneller Unterricht
Versuchsphase I "Radizieren"	1 Gruppe B "knowledge" "comprehension"	2 Gruppe A "knowledge" "comprehension"
Versuchsphase II "Logarithmieren"	3 Gruppe A "knowledge" "comprehension"	4 Gruppe B "knowledge" "comprehension"

Der Versuchsplan ermöglicht folgende Vergleiche:

- Vergleich ein und derselben Stichprobe, welche zwei Unterrichtsmethoden ausgesetzt wurde (Feld 1 versus 4 und Feld 3 versus 2).
- Vergleich zweier unabhängiger Stichproben, welche je einer Behandlungsmethode ausgesetzt wurden (Feld 1 versus Feld 2 und Feld 3 versus Feld 4).

Untersuchungsrichtung a)

Diese Vergleichsrichtung bietet den großen Vorteil, daß der Stichprobenfehler reduziert ist, da hier eine wiederholte Messung an einer Stichprobe vorliegt. Voraussetzung für einen solchen Vergleich ist allerdings, daß die bei den beiden Versuchsdurchgängen verwendeten Meßinstrumente aufeinander bezogen werden können und "dieselbe Leistung" gemessen wird¹⁾. Solange dies nicht der Fall ist, können unterschiedliche Effekte auf formale und inhaltliche Divergenzen der verwendeten Meßinstrumente zurückgeführt werden. In diesem Zusammenhang kann im vorhinein vermerkt werden, daß im Rahmen unserer Untersuchung eine Transformation von Rohwerten der Testergebnisse der Zellen 1 und 2 bzw. 3 und 4 in Standardwerte nicht möglich war, da wir aufgrund der verschiedenen Behandlungen annehmen müssen, daß die Maßzahlen der Felder 1 und 2 bzw. 3 und 4 nicht derselben Population angehören. Wir haben daher, um eine Vergleichbarkeit der Maßzahlen der Tests zu gewährleisten, die Tests "knowledge-Radizieren" und "knowledge-Logarithmieren" bzw. die Tests "comprehension-Radizie-

1) Das Problem, Maßzahlen unmittelbar miteinander zu vergleichen, welche nicht unter denselben Bedingungen gewonnen wurden, stellt sich in ähnlicher Weise bei dem Vergleich abhängiger Gruppenprofile (vgl. G.A. LIENERT: "Eine varianzanalytische Methode zum Nachweis der Gruppenspezifität von Testprofilen", in: Zeitschrift f. exp. u. angew. Psych., 10, 1963, S. 333-345). Da die zum Vergleich solchermaßen gewonnener Maßzahlen notwendigen Bedingungen trotz unserer im Rahmen der Testkonstruktion getroffenen Vorkehrungen im strengen Sinne niemals zu erfüllen sind und daher systematische Ergebnisverzerrungen nicht ausgeschlossen werden können, werden wir die von uns behauptete höhere Lehreffektivität der programmierten Unterweisung erst dann als gegeben ansehen, wenn sie bei verschiedenen Vergleichsrichtungen signifikant abgesichert ist.

ren" und "comprehension-Logarithmieren" so zusammengestellt, daß sie bezüglich der mittleren Schwierigkeitsindices und damit der arithmetischen Mittel der Rohwerte übereinstimmen¹⁾. Auf alle damit zusammenhängenden Probleme sowie die entsprechenden statistischen Kennwerte der Tests werden wir unter Punkt 7.81 ausführlich eingehen. Die inhaltliche Vergleichbarkeit der gemessenen Leistungen scheint uns dadurch gewährleistet zu sein, daß wir mit den Unterrichtsstoffen "Radizieren" bzw. "Logarithmieren" jeweils eine Umkehrung des "Potenzierens" gewählt haben. Die Begriffe "Basis", "Exponent" und "Potenzwert" im Potenzrechnen entsprechen den Begriffen "Wurzelwert", "Wurzelexponent" und "Radikant" beim Radizieren bzw. den Begriffen "Basis", "Logarithmus" und "Numerus" beim Logarithmieren²⁾. Unsere Prüfungen messen also nicht nur "algebraische Rechenfertigkeit" und "algebraisches Verständnis", darüber hinaus wird das weitgesteckte Feld der Algebra auf Rechentechniken eingeschränkt, welche auf strukturverwandten bzw. strukturgleichen (das ist eine ausschließliche Frage der Definition!) mathematischen Operationen basieren. Trotz dieser weitgehenden Übereinstimmung handelt es sich um verschiedene Schreibweisen und Rechentechniken, so daß nicht von einem

1) Bei der Streuung der Schwierigkeitsindices ist uns dies nur insofern gelungen, als wir feststellen können, daß sich die Streuungen der aufeinander bezogenen Tests nicht signifikant unterscheiden (vgl. 8.10 e)).

2) Vg. H. FLEBIG und H. JUNGHANS: "Rechnen mit Logarithmen", Braunschweig o.J., S. 11

d i r e k t e n Übungseffekt gesprochen werden kann¹⁾. So kann z.B. ein Schüler perfekt gelernt haben, Wurzel-
ausdrücke zu multiplizieren, ohne in der Lage zu sein,
selbst einfachste Multiplikationsaufgaben mit logarith-
mischen Ausdrücken durchzuführen. Ein indirekter Übungs-
gewinn muß trotzdem angesetzt werden, er besteht vor
allem darin, daß im Verlauf des Rechnens mit Wurzeln und
mit Logarithmen im Potenzrechnen eingeübte Fertigkeiten
miteingehen²⁾. Wir werden auf dieses Problem bei der
Analyse möglicher Ergebnisverzerrungen durch Übungs-
effekte und Lernübertragung noch ausführlich zu spre-
chen kommen. Zusammenfassend können wir festhalten, daß
wir durch die Konstruktion von Tests, deren Maßzahlen
ohne Transformation und Korrekturgrößen direkt miteinan-
der verglichen werden können, sowie durch die Wahl von

-
- 1) Eine formale Entsprechung scheinen uns Subtests im
Rahmen von Intelligenztests darzustellen, welche
sich inhaltlich deutlich unterscheiden und trotzdem
verschiedene Aspekte ein und desselben Intelligenz-
faktors messen. So messen z.B. die Subtests 7 und 8
im Leistungsprüfsystem von HORN beide den Intelli-
genzfaktor "Technische Begabung", wobei im Subtest
7 von der Versuchsperson verlangt wird, aus einer
vorgegebenen Anzahl an Buchstaben bzw. Zahlen die-
jenigen anzustreichen, welche "räumlich verkehrt" dar-
gestellt sind, im Subtest 8 müssen dagegen Kanten
und Flächen eines Grundrisses den entsprechenden
Stellen eines dazugehörigen Körpers zugeordnet wer-
den. Auch hier kann davon ausgegangen werden, daß
durch die Bearbeitung des Subtests 7 für die Bearbei-
tung des Subtests 8 kein d i r e k t e r Übungs-
gewinn entsteht (vgl. W. HORN: "Leistungsprüfsystem
L-P-S", Göttingen 1962).
 - 2) Aus diesem Grund hielten wir es für unumgänglich,
unsere Versuchsgruppen bezüglich der Kenntnisse im
Potenzrechnen aneinander anzugleichen.

Unterrichtsstoffen, welche dieselben "arithmetischen Fähigkeiten" voraussetzen, die Voraussetzung dafür geschaffen haben, daß wir die Maßzahlen der Felder 1 und 4 bzw. 3 und 2 miteinander vergleichen und auf die verschiedene Einwirkung unserer Unterrichtsmethoden hin analysieren können. Dabei haben wir durch die oben geschilderten Vorkehrungen weitgehend verhindert, daß unterschiedliche Effekte auf Abweichungen der Meßinstrumente bzw. der behandelten Stoffgebiete zurückzuführen sind. Da trotz unserer Vorkehrungen Wechselwirkungseffekte, welche den Einfluß unserer Unterrichtsmethoden verfälschen, auftreten können, werden wir die von uns hypothetisch angenommene höhere Lehreffektivität der Programmierten Unterweisung erst dann als bestätigt ansehen, wenn sie mindestens doppelt signifikant abgesichert ist, d.h. also sowohl bei Vergleich der Felder 1 und 4 als auch der Felder 3 und 2.

Untersuchungsrichtung b)

Die Wahrscheinlichkeit, daß bei der Untersuchungsrichtung a) Wechselwirkungseffekte sich dahingehend verfälschend auswirken, daß sowohl bei Vergleich der Felder 1 versus 4 als auch der Felder 3 versus 2 eine signifikant höhere Lehreffektivität der Programmierten Unterweisung auftritt, welche nur scheinbar durch die Unterrichtsmethoden verursacht wurde, ist angesichts der getroffenen Vorkehrungen weitaus geringer als die Wirkung von Störfaktoren, mit welcher bei der einfachen quasi-experimentellen Anordnung mit nicht gleichartiger Kontrollgruppe gerechnet werden muß. Um diese Wahrscheinlichkeit noch weiter zu reduzieren, haben wir zusätzlich den Vergleich der Felder 1 versus 2 bzw. der Felder 3 versus 4 angesetzt. Für sich allein betrachtet

handelt es sich bei jeder dieser Vergleichsrichtungen um eine quasiexperimentelle Anordnung mit nicht gleichartiger Kontrollgruppe. Selbst bei einem Verzicht auf die Untersuchungsrichtung a) würde sich hier unser Vorgehen vorteilhaft von den üblichen einfachen Versuchsanordnungen abheben, da wir die Fragerichtung wiederholen, d.h. jede Gruppe jeder Behandlung aussetzen. Sollte es uns gelingen, auch bei diesen beiden Vergleichsrichtungen, welche nicht mit der Problematik der nur weitgehend formal und inhaltlich aneinander angepaßten Testinstrumente behaftet sind¹⁾, eine signifikant höhere Lehreffektivität der programmierten Unterweisung nachzuweisen, wäre es uns mit noch höherer Wahrscheinlichkeit gelungen, die Möglichkeit auszuschließen, daß die gefundenen unterschiedlichen Gruppenleistungen auf den Einfluß von Störvariablen zurückgeführt werden können, d.h. nur von vorgetäuschten Methodeneffekten bewirkt sind. In diesem Fall könnten wir mit einer sehr hohen Zuverlässigkeit unserer Ergebnisse rechnen.

Bei der weiteren Diskussion unseres Versuchsplans müssen wir noch auf den möglicherweise verfälschenden Effekt eingehen, der auf Übung und Lernübertragung zurückgeführt werden kann. Wir haben bereits oben erwähnt, daß aufgrund des gemeinsamen Bezugs der beiden angesetzten Rechentechniken zum Potenzrechnen ein Übungsfortschritt und eine Lernübertragung durch die Absolvierung der Versuchsphase I verursacht werden kann, welche die Ein-

1) Bei dieser Vergleichsrichtung ist allerdings die Wahrscheinlichkeit, daß durch den größeren Stichprobenfehler tatsächlich vorhandene unterschiedliche Methodeneffekte verdeckt werden, ziemlich hoch. Dies gilt umso mehr, als wir ja keine Zufallsstichproben vergleichen.

gangsbedingungen der Versuchspersonen in die Versuchsphase II gegenüber denen der Versuchsphase I verändern. Hier muß einmal ein genereller, für alle Gruppen gleichermaßen geltender Effekt angesetzt werden. Zum anderen kann aufgrund unserer Vermutung, daß die Programmierte Unterweisung eine höhere Lehreffektivität bewirkt als der herkömmliche Unterricht, angenommen werden, daß die Gruppe B im Verlauf der Versuchsphase I einen höheren Lernfortschritt erzielt als die Gruppe A. Daraus folgt nun, daß die beiden Gruppen nicht mehr unter denselben Voraussetzungen in die Versuchsphase II eintreten, vielmehr muß (bei Annahme einer Lern- und Übungsübertragung) davon ausgegangen werden, daß die Gruppe B eine bessere Ausgangsposition hat, d.h. daß die Lernübertragung nicht für beide Gruppen gleichmäßig erfolgte. Da wir nicht von vornherein die Wirksamkeit dieser Störvariable als unbedeutende systematische Fehlervarianz vernachlässigen können, müssen wir an dieser Stelle die theoretisch zu erwartenden Auswirkungen auf unsere Untersuchungsergebnisse analysieren. Dabei müssen wir bei dieser Analyse von der von uns hypothetisch gesetzten Überlegenheit der Programmierten Unterweisung ausgehen. Beim Vergleich der Felder 1 versus 4 wird der von uns erwartete Unterschied nivelliert, d.h. die angenommene Störvarianz wirkt g e g e n - l ä u f i g zur Richtung unserer Hypothese. Beim Vergleich der Felder 3 versus 2 ergibt sich folgendes Bild: Die Gruppe A geht - im Fall der höheren Lehreffektivität der Programmierten Unterweisung - zwar mit schlechteren Lernvoraussetzungen in die Untersuchungsphase II als die Gruppe B, dies schließt allerdings einen grundsätzlichen Übungsgewinn durch die Absolvierung der Lernphase I nicht aus, so daß sich hier die Störvarianz in Richtung unserer Hypothese auswirkt, wenngleich dies

nicht in dem Umfang der Fall sein dürfte wie umgekehrt im Fall des Vergleichs von Feld 1 versus Feld 4. Da wir jedoch angesichts der von uns vermuteten Überlegenheit der Programmierten Unterweisung davon ausgehen müssen, daß die Gruppe B den zweiten Lerndurchgang mit besseren Lernvoraussetzungen angeht als die Gruppe A, müssen wir gleichzeitig annehmen, daß im Falle des Vergleichs von Feld 3 mit Feld 4 die systematische Störvarianz der Richtung unserer Hypothese entgegenwirkt¹⁾.

Schließlich muß noch auf Störeffekte hingewiesen werden, welche aufgrund der Behandlungsabfolge verursacht werden und Methodeneffekte vortäuschen bzw. nivellieren können, wobei streng genommen nur ersteres von Bedeutung ist. Es war ursprünglich geplant, diese Möglichkeit der Verfälschung unserer Versuchsergebnisse dadurch unter Kontrolle zu halten, daß die Abfolge der Behandlungen systematisch variiert wird. Diese Absicht mußte jedoch aufgrund organisatorischer Schwierigkeiten aufgegeben werden²⁾. Ungeachtet der Tatsache, daß durch den Verzicht auf die systematische Variation der Behandlungsabfolge eine mögliche systematische Störvarianz unkontrolliert bleibt, meinen wir festhalten zu können, daß eine nennenswerte Beeinflussung der abhängigen Variablen durch die Behandlungsfolge ziemlich unwahrscheinlich ist.³⁾

Zum Schluß sei noch kurz auf die Ausgangsbedingungen der beiden Gruppen bezüglich der Kriterien "Intelligenz" und "Vorwissen" eingegangen (die entsprechenden Zahlenangaben finden sich unter Punkt 7.50). Wir haben schon mehrfach erwähnt, daß durch die Feststellung nichtsignifikanter Unterschiede bezüglich solcher Kriteriumsvariablen zwischen den zur Untersuchung herangezogenen Gruppen keinesfalls mit der nötigen Sicherheit behauptet werden kann, daß die Stichproben vergleichbar sind. Wenn wir trotzdem diese Ausgangsbe-

- 1) Selbst wenn sich beim Vergleich der Felder 3 versus 4 ein nicht signifikanter Unterschied ergeben sollte, können wir, sofern bei allen übrigen Vergleichsrichtungen signifikante Unterschiede auftreten, mit hoher Wahrscheinlichkeit davon ausgehen, daß die Programmierte Unterweisung eine signifikant höhere Lehreffektivität aufweist als der herkömmliche Unterricht.
- 2) Wir beabsichtigten ursprünglich insgesamt 8 Klassen zu unserer Untersuchung heranzuziehen. Im Verlauf der vorexperimentellen Phase stellte sich allerdings heraus, daß zwei Klassen im Wurzelrechnen bereits einschlägige Vorkenntnisse aufwiesen, da in diesen die Quadratwurzel bereits behandelt war. Es wäre zwar grundsätzlich möglich gewesen, die verbleibenden 6 Klassen (mehr waren in Bayreuth nicht verfügbar) so aufzuteilen, daß eine systematische Variation der Behandlungsfolge möglich gewesen wäre. In diesem Fall wären wir allerdings - um nur eine Fehlerquelle aufzuzeigen - aufgrund der vorgegebenen Stundenpläne, an welche wir uns halten mußten, gezwungen gewesen, eine Zelle mit nur 16 Versuchspersonen zu besetzen. Angesichts der z.T. recht umfangreichen experimentellen Einbußen, mit welchen im Rahmen von einschlägigen Untersuchungen gerechnet werden muß und die nur unzureichend "vorausberechnet" werden können, schien uns das Risiko, welches mit einer Beibehaltung unseres ursprünglichen Versuchsplans verbunden gewesen wäre, zu groß zu sein. Der Rückgriff auf Klassen außerhalb von Bayreuth war aus organisatorischen Gründen nicht möglich, da wir über die Beobachter nicht beliebig verfügen konnten.
- 3) Seitens einiger befragter Mathematiklehrer wurde uns übereinstimmend versichert, daß die Abfolge der beiden Stoffgebiete beliebig variierbar sei.

dingungen ermittelt und durch Eliminierung von Versuchspersonen manipuliert haben, so geschah dies deswegen, weil die entsprechenden Daten ergänzend zur Diskussion unseres Versuchsplans und der Versuchsergebnisse herangezogen werden können. Darüber hinaus muß darauf hingewiesen werden, daß auch ein Kreuzvalidierungskonzept mit umso weniger Erfolg herangezogen werden kann, je verschiedener die Ausgangsbedingungen der Gruppen bezüglich der für die Lernleistung so zentralen Kriterien wie "Intelligenz" und "Vorwissen" sind. Dies wollen wir uns an einem Beispiel kurz vergegenwärtigen: Es ist durchaus möglich, daß Gruppen verschiedener Intelligenz auf unsere Unterrichtsmethoden verschieden ansprechen. So wäre es z.B. denkbar und wird auch öfters behauptet¹⁾, daß eine weniger intelligente Gruppe im Gegensatz zu einer intelligenten Gruppe mit programmiertem Arbeitsmaterial signifikant besser arbeitet als im herkömmlichen Vergleichsunterricht. Wenn solchermaßen divergierende Gruppen in ein Kreuzvalidierungskonzept aufgenommen werden, kann es geschehen, daß zwar in einer Vergleichsrichtung eine signifikante Überlegenheit der Programmierten Unterweisung festgestellt werden kann, im anderen Fall jedoch nicht²⁾. Die solchermaßen gewonnenen Ergebnisse hätten nicht mehr Aussagekraft als diejenigen

1) Vgl. Punkt 2.30

2) Die divergierenden Ergebnisse der beiden Vergleiche wären in diesem Fall auf die ungleiche Stichprobensammensetzung zurückzuführen.

einer quasiexperimentellen Anordnung mit nicht gleichartiger Kontrollgruppe, da die Überlegenheit der ausbalancierten Versuchsanordnung u.a. darin besteht, daß für a l l e behandelten Gruppen die höhere Effektivität einer Unterrichtsmethode aufgezeigt werden kann. Aus diesem Grund haben wir unsere beiden Gruppen im Hinblick auf die Kriterien "Intelligenz" und "Vorwissen" analysiert und konnten zu der Feststellung gelangen, daß sie sich diesbezüglich nichtsignifikant unterscheiden. Durch diese Bedingungen erhöhte sich die Wahrscheinlichkeit, daß wir im Rahmen unseres Kreuzvalidierungskonzepts zu Ergebnissen gelangen konnten, welche a l l e in derselben Richtung liegen. Im Zusammenhang mit dem Kriterium "Vorwissen" weisen wir vorwegnehmend ausdrücklich darauf hin, daß bezüglich der behandelten Stoffe keinerlei direkte Vorkenntnisse vorlagen, da wir alle Schüler, welche in der Lage waren, selbst e i n f a c h e Aufgaben im "Wurzelrechnen" und "Rechnen mit Logarithmen" zu lösen, zwar in der Untersuchung mitlaufen ließen, bei der endgültigen Verrechnung jedoch nicht berücksichtigten. Darüber hinaus wurden alle diejenigen Schüler in die verrechnete Stichprobe nicht aufgenommen, welche sich bezüglich ihrer Vorkenntnisse im Potenzrechnen extrem von der Verteilung der Maßzahlen aller Gruppenmitglieder abhoben und dadurch die Vergleichbarkeit der beiden Gruppen beeinträchtigt hätten¹⁾. Bei der Kriteriumsvariable "Intelligenz" waren solche Vorkehrungen nicht notwendig.

1) Ein entsprechendes Vorgehen findet sich u.a. bei STRITTMATTER (vgl. P. STRITTMATTER, a.a.O., S. 179), welcher von seinen 6 Versuchsklassen 20 % der Schüler nicht zur Auswertung mit heranzog, da sie sich bezüglich der Kriterien "Intelligenz" und "Schulleistung in einem untersuchungsbezogenen Fach" nicht in die gebildeten Tripels einfügen ließen.

7.40 Versuchsablauf

7.41 Vorexperimentelle Phase

Im Verlauf der vorexperimentellen Phase übten die Beobachter das Protokollieren der Beobachtungskategorien (vgl. 7.72). Dadurch wurden die Schüler an das Vorhandensein der drei Beobachter gewöhnt. Insgesamt wurden dafür 3 bzw. 2 Stunden pro Klasse¹⁾ aufgewendet. Weiter wurden die adressatenbezogenen Maßkriterien "Intelligenz", "Einstellung zur Schule" und "Einstellung zum Lehrer" ermittelt. Dafür waren insgesamt 3 Stunden pro Klasse erforderlich. Unmittelbar vor Beginn der experimentellen Phase wurden die Vorkenntnisse im "Potenzrechnen" geprüft (1 Stunde pro Klasse). Die an der Untersuchung beteiligten Lehrer waren aufgefordert worden, ihre Klassen im Potenzrechnen auf einen vergleichbaren Wissensstand zu bringen. Die dafür notwendigen Lehrziele waren ihnen rechtzeitig angegeben worden. Weiter wurde im Verlauf einer Stunde die Arbeit mit einem Unterrichtsprogramm geübt²⁾. Keine der Klassen war bis zu diesem Zeitpunkt mit Unterrichtsprogrammen in Berührung gekommen.

-
- 1) Die Gewöhnungsphase dauerte lediglich 2 Stunden pro Klasse. Anschließend wurden allerdings in vier der sechs Versuchsklassen eine weitere Stunde protokolliert, um die für eine Reliabilitätsprüfung der Beobachtungskategorien notwendigen Daten zu erhalten.
 - 2) Dazu wurde ein Programm verwendet, welches bezüglich seines formalen Aufbaus mit den in der experimentellen Phase verwendeten Programmen übereinstimmt. (H. FIEBIG und H. JUNGHANS: "Rechnen mit Potenzen", Braunschweig o.J.)

7.42 Experimentelle Phase

Aufgrund der komplizierten Stundenplanorganisation war es nicht möglich, für alle Klassen absolut identische zeitliche Bedingungen zu schaffen. Der Versuchsunterricht wurde jedoch grundsätzlich vormittags durchgeführt, pro Tag wurde für jede Klasse grundsätzlich nur e i n e Unterrichtsstunde eingeplant. Die Versuchsunterrichte fanden in den Klassenzimmern der Versuchsklassen statt und wurden im Rahmen der im Stundenplan angesetzten Mathematikstunden abgehalten. Die Versuchsphase I lief über 5 Schulstunden¹⁾. Anschließend wurden 3 Prüfungen durchgeführt, von denen zwei über je 45 Minuten liefen und eine über 20 Minuten (pro item zirka zwei Minuten). Die Prüfungen wurden auf aufeinanderfolgende Tage verteilt. Unmittelbar danach begann die Versuchsphase II, sie erstreckte sich ebenfalls über 5 Schulstunden. Danach wurden 2 Prüfungen à 45 Minuten angesetzt (pro item zirka zwei Minuten). Nach Abschluß des Versuchsunterrichts wurde die Einstellung zum programmierten bzw. zum konventionellen Unterricht gemessen.

1) Eine Schulstunde umfaßt lediglich 45 Minuten. Da wir darauf Wert legten, daß jeweils wirklich 45 Minuten unterrichtet wurde, wurden die Lehrer aufgefordert, organisatorische Dinge nicht im Rahmen dieser Versuchsstunden abzuhandeln. Der Unterrichtsbeginn wurde von einem Beobachter registriert. Dieser gab zwei Minuten vor Ablauf der 45 Minuten dem Lehrer ein Zeichen, dieses wiederholte er am Ende der Stunde. Wurde die Zeit trotzdem überzogen, wurde dies in der darauffolgenden Unterrichtsstunde berücksichtigt. Ein solches Vorgehen schien uns angebracht zu sein, da wir ein abruptes Abbrechen der Versuchsstunden vermeiden wollten um die Reaktivität der experimentellen Situation einzuschränken.

Die Absicht, nach Ablauf von einem Monat eine Behaltensprüfung durchzuführen, scheiterte am Lehrkonzept der an der Untersuchung beteiligten Lehrer¹⁾.

7.43 Instruktionen

Die Schüler wurden darüber informiert, daß die Verwendungsfähigkeit einer neuen Unterrichtsmethode untersucht würde. Es wurde ihnen allerdings kein Hinweis dahingehend gegeben, daß hypothetisch eine Überlegenheit der Programmierten Unterweisung angenommen wurde. Die Schüler wurden weiter darauf hingewiesen, daß am Ende jeder Versuchsphase Prüfungen angesetzt seien und daß die dort erbrachte Leistung für die Zeugnisgebung im Sinne einer üblichen Schularbeit herangezogen werde. Weitere Motivierungen wurden nicht gegeben. Bezüglich der adressatenbezogenen Maßkriterien wurde ihnen mitgeteilt, daß die Auswertung anonym erfolge und die Lehrer in keinem Fall Einsicht in diese Unterlagen bekämen.

1) Die Lehrer bestanden darauf, unmittelbar im Anschluß an den Versuchsunterricht einen weiterführenden und vertiefenden Unterricht der behandelten Stoffgebiete durchzuführen, da sie im Falle einer einmonatigen Pause sehr viel Zeit für die Wiederauffrischung des Gelernten aufbringen müßten. Angesichts der Tatsache, daß sie bereits durch die Untersuchung einige Unterrichtsstunden "verloren" und im Verlauf des Schuljahres noch ein sehr großes Pensum zu bewältigen hatten, war ihr Ansinnen verständlich. Damit entfiel allerdings die für die Gedächtnisprüfung notwendige Voraussetzung, daß zwischenzeitliche systematische Lerneinflüsse bezüglich der geprüften Stoffgebiete vorgelegen hätten. Angesichts der Tatsache, daß sich fast grundsätzlich bei vergleichbaren Untersuchungen keine wesentlichen Verschiebungen der Befunde zwischen unmittelbarer Prüfung und der Gedächtnisprüfung ergaben, scheint uns hier keine wesentliche Information entgangen zu sein.

Weiter waren die Schüler grob darüber informiert, daß die Beobachter unterrichtliche Aktivitäten protokollieren. Diese Information schien uns aufgrund des relativ auffälligen organisatorischen Aufwandes notwendig zu sein. Insbesondere legten wir Wert auf die Feststellung, daß nicht nur die unterrichtlichen Aktivitäten der Schüler sondern auch die des Lehrers festgehalten würden. Eine detaillierte Information über den Inhalt der Beobachtungskategorien wurde nicht gegeben. Die Lehrer waren über die globale Zielsetzung des Experiments genau informiert, sie wurden ausdrücklich aufgefordert, einen möglichst effektiven Unterricht zu halten, damit die Voraussetzungen für eine strenge Vergleichsbasis geschaffen würden. Allerdings wurde sorgfältig vermieden, detaillierte Informationen über den Inhalt der Beobachtungskategorien zu geben, um jede Beeinflussung der unterrichtlichen Lehreraktivitäten zu vermeiden. Die für die vergleichbare Gestaltung des herkömmlichen Unterrichts notwendigen Informationen finden sich unter Punkt 7.71. Die Lehrer waren schließlich darüber informiert, daß während der Dauer des Experiments keine zusätzlichen Mathematikstunden stattfinden durften.

7.50 Untersuchte Stichprobe

Die Untersuchung wurde in sechs Klassen des 10. Schuljahrgangs in Bayreuther Gymnasien durchgeführt. Der ursprüngliche Stichprobenumfang betrug 149 Versuchspersonen, davon waren 90 weiblich und 59 männlich. Im Verlauf der Untersuchung reduzierte sich die Stichprobe

auf 119 Schüler. Einmal mußte bei der Endauswertung auf alle diejenigen Schüler verzichtet werden, welche die Klasse wiederholten, da diesen die in der Untersuchung verwendeten Unterrichtsstoffe "Wurzelrechnen" und "Rechnen mit Logarithmen" bereits bekannt waren. Zum anderen konnten nur diejenigen Schüler in die Endauswertung aufgenommen werden, welche während der gesamten Untersuchungsdauer regelmäßig anwesend waren. Schließlich wurden alle Schüler, welche bezüglich des durchschnittlichen Vorwissens *e x t r e m* von der Verteilung der Maßzahlen der jeweiligen Gruppe abwichen, in die zur Verrechnung herangezogene Stichprobe nicht aufgenommen. Die Zuordnung der sechs Klassen zu den beiden Versuchsgruppen mußte im Hinblick auf die Klassenstärke und den jeweils vorhandenen Stundenplan vorgenommen werden. Letzteres war unumgänglich, da die Gymnasien aufgrund des Fachlehrersystems eine komplizierte Stundenplanorganisation aufwiesen, welche eine willkürliche Festlegung der für die Untersuchung benötigten Stunden nicht zuließ.

Die beiden Versuchsgruppen setzten sich ursprünglich wie folgt zusammen:

Gruppe A	Schule	Gruppe B	Schule
Klasse 1: 31 Schüler (gemischt)	a	Klasse 4: 23 Schüler (männlich)	d
Klasse 2: 16 Schüler (gemischt)	b	Klasse 5: 20 Schüler (weiblich)	c
Klasse 3: 30 Schüler (weiblich)	c	Klasse 6: 29 Schüler (gemischt)	b

77 Schüler

(58 weiblich
19 männlich)

72 Schüler

(32 weiblich
40 männlich)

In der endgültigen, zur Verrechnung herangezogenen Stichprobe ergab sich folgende Zusammensetzung:

Gruppe A

Gruppe B

Klasse 1: 20 Schüler	Klasse 4: 13 Schüler
Klasse 2: 14 Schüler	Klasse 5: 19 Schüler
Klasse 3: 28 Schüler	Klasse 6: 25 Schüler

N = 62

N = 57

(47 weiblich
15 männlich)

(27 weiblich
30 männlich)

Durchschnittsalter: 15,74

Durchschnittsalter: 15,71

Altersstreuung: 15-17

Altersstreuung: 15-17

Beide Versuchsgruppen wurden hinsichtlich "Intelligenz"¹⁾ und "Vorwissen"²⁾ miteinander verglichen³⁾. Dabei ergaben sich folgende Werte⁴⁾:

-
- 1) Zur Messung der Intelligenz wurde der Intelligenz-Struktur-Test von AMTHAUER herangezogen (R. AMTHAUER: "Intelligenz-Struktur-Test", Göttingen 1955)
 - 2) Das Vorwissen wurde mit einem von uns zusammengestellten kriteriumsbezogenen Test gemessen. Eine ausführliche Diskussion dieses Verfahrens findet sich unter 7.81 .
 - 3) Die zur Verrechnung verwendeten Formeln finden sich in E. MITTENECKER: "Planung und statistische Auswertung von Experimenten", Wien 1963, S. 59 f. und S. 69 .
 - 4) Diese Werte beziehen sich auf die endgültig verrechneten Stichproben.

a) Gesamtintelligenz¹⁾:

Gruppe A	Gruppe B
$M_A \approx 117,08$	$M_B \approx 115,86$
$s^2_A \approx 73,66$	$s^2_B \approx 83,38$

Differenz der Mittelwerte: $\sigma_{\text{Diff}} \approx 1,622$

$t \approx 0,752$

($t_{\text{krit}} 5\% \approx 1,98$)

(nicht sign.)

Vergleich der Varianzen: $F \approx 1,13$

(nicht sign.)

b) "Mathematische Intelligenz"²⁾:

Gruppe A	Gruppe B
$M_A \approx 107,34$	$M_B \approx 105,88$
$s^2_A \approx 68,10$	$s^2_B \approx 62,23$

Differenz der Mittelwerte: $\sigma_{\text{Diff}} \approx 1,480$

$t \approx 0,986$

($t_{\text{krit}} 5\% \approx 1,98$)

(nicht sign.)

Vergleich der Varianzen: $F \approx 1,09$

(nicht sign.)

1) Ausgedrückt in Standardwerten (SW)

2) Da aufgrund der verschiedenen Schulzugehörigkeit vermutet werden kann, daß bezüglich der "mathematischen Komponente" der Intelligenz systematische Unterschiede zwischen unseren beiden Gruppen A und B bestehen, haben wir zusätzlich zu dem Wert SW-Gesamt das arithmetische Mittel der Subtests 6 (ZR: "Induktives Denken-können mit Zahlen") und 7 (RA: "Rechnerisches Denken, Arithmetik") zur Prüfung der Mittelwertsdifferenz herangezogen.

c) Vorwissen¹⁾:

Gruppe A	Gruppe B
$M_A \approx 6,05$	$M_B \approx 6,09$
$s_A^2 \approx 3,36$	$s_B^2 \approx 4,13$

Differenz der Mittelwerte: $\delta_{\text{Diff}} \approx 0,347$

$t \approx 0,115$
 $(t_{\text{krit}} 5\% \approx 1,98)$
 (nicht sign.)

Vergleich der Varianzen: $F \approx 1,23$
 (nicht sign.)

Die Zuteilung der Schulen zu dem Versuch erfolgte nicht auf der Basis einer freiwilligen Zusage, sondern durch Zuordnung seitens des Ministerialbeauftragten für Gymnasien in Oberfranken. Ungeachtet dieser Tatsache muß die Bereitschaft der an der Untersuchung beteiligten Lehrer als gut bezeichnet werden. Die Klassen 3 und 5 wurden von derselben Lehrerin unterrichtet. Abgesehen von der Klasse 5 waren alle an der Untersuchung beteiligten Lehrer gleichzeitig Klassenlehrer in den Versuchsklassen. Alle an der Untersuchung beteiligten Lehrer waren voll ausgebildete Fachlehrer für Mathematik und wiesen eine langjährige Berufspraxis an der jeweiligen Schule auf. Die Wochenstundenzahl für das Fach Mathematik war mit Ausnahme der Schule d überall gleich, in der Schule d wurde wöchentlich eine Stunde mehr im Fach Mathematik

1) Die Verteilungsform der Maßzahlen wurde überprüft. Dabei ergab sich bei beiden Gruppen keine signifikante Abweichung von der Normalverteilung.

unterrichtet. Alle Klassen hatten keinerlei spezielle Vorkenntnisse im "Wurzelrechnen" und "Rechnen mit Logarithmen". Im Potenzrechnen wiesen sie bezüglich des behandelten Stoffes einen vergleichbaren Wissensbestand auf.

7.60 Verwendete Unterrichtsprogramme

Sowohl das in der vorexperimentellen Phase zur Eingewöhnung verwendete Programm als auch die eigentlichen Versuchsprogramme stammen von denselben Autoren. Es handelt sich dabei um folgende Programme:

- "Rechnen mit Potenzen"¹⁾
- "Rechnen mit Wurzeln"²⁾
- "Rechnen mit Logarithmen"³⁾

Die Programme sind nach demselben Konstruktionsprinzip entworfen und gleichen sich daher bezüglich des formalen Aufbaus vollkommen. Es handelt sich um Linearprogramme. Am Ende jedes Lernschrittes stehen eine oder mehrere Aufgaben, welche vom Schüler selbständig schriftlich bearbeitet werden müssen. Die Lösung wird in das beigelegte Antwortheft geschrieben. Auf der nachfolgenden Seite des Programms steht die Lösung der vom Schüler bearbeiteten Aufgabe mit einer auf die umseitige Information bezogenen Erklärung des Lösungsweges. Diese

1) H. FIEBIG und H. JUNGHANS: "Rechnen mit Potenzen", Westermann-Verlag, o.J.

2) H. FIEBIG und H. JUNGHANS: "Rechnen mit Wurzeln", Westermann-Verlag, o.J.

3) H. FIEBIG und H. JUNGHANS: "Rechnen mit Logarithmen", Westermann-Verlag, o.J.

zusätzliche Erklärung ist vom übrigen Text farbig abgehoben und muß von den Schülern nicht durchgelesen werden. Der Stoff ist nach sachlogisch begründeten Abschnitten in einzelne Programmeinheiten gegliedert. Am Ende jeder Programmeinheit findet sich eine aus 10 Fragen bestehende Wissensprüfung, deren Lösungen auf einem dem Programm beigelegten Lösungsblatt ohne weitere Erklärung für den Schüler jederzeit eingesehen werden können. Von dem im ersten Versuchsdurchgang angesetzten Programm "Rechnen mit Wurzeln" wurden die drei ersten Programmeinheiten bearbeitet.

Es handelt sich dabei um:

- Programmeinheit 1: Wurzelbegriff,
Berechnung der Quadrat- 16 Lernschritte
wurzel
- Programmeinheit 2: Bruchpotenzen 14 Lernschritte
- Programmeinheit 3: Addition, Subtraktion,
Multiplikation und 16 Lernschritte
Division von Wurzeln

Von dem im zweiten Versuchsdurchgang angesetzten Programm "Rechnen mit Logarithmen" wurden ebenfalls die ersten drei Programmeinheiten bearbeitet:

- Programmeinheit 1: Logarithmusbegriff 14 Lernschritte
- Programmeinheit 2: Rechnen mit Logarithmen:
Multiplikation und 14 Lernschritte
Division
- Programmeinheit 3: Rechnen mit Logarithmen:
Potenzieren und Radizieren 12 Lernschritte

Die Programme wurden in Buchform geboten. Jeder Schüler hatte sein eigenes Programm¹⁾. Die zur Bearbeitung der von uns gewählten Programmeinheiten notwendige Zeit betrug 5 Schulstunden. Sie wurde nach Rücksprache mit den Programmautoren, welche bereits einschlägige Erfahrungen gesammelt hatten, festgelegt. Dabei gingen wir davon aus, daß auch die schwachen Schüler genügend Zeit haben sollten, alle Lernschritte durchzuarbeiten²⁾. Schüler, welche vor Ablauf der fünf Stunden die vorgesehenen Lernschritte durchgearbeitet hatten, erhielten den Auftrag, den Lernstoff zu wiederholen.

7.70 Der "herkömmliche Vergleichsunterricht"

Zusätzlich zur Definition der Makrostruktur unseres "herkömmlichen Vergleichsunterrichts", welche wir unter Punkt 4.41 ausführlich geschildert haben, wurden die Lehrer von uns mit detaillierten Anweisungen versehen, an welche sie sich bei der praktischen Durchführung des Unterrichts zu halten hatten. Dies wurde von uns als Sicherung der "Mikrostruktur" des "herkömmlichen Vergleichsunterrichts" bezeichnet. Weiter hielten wir mittels einer sorgfältigen Protokollierung

-
- 1) Die zur Durchführung der Untersuchung notwendigen Klassensätze wurden vom Georg-Westermann-Verlag unentgeltlich zur Verfügung gestellt.
 - 2) Anhand der Antworthefte konnten wir feststellen, daß lediglich zwei Schüler nach Ablauf der 5 Schulstunden die vorgesehenen Lernschritte noch nicht bewältigt hatten. Sie wurden zur endgültigen Auswertung nicht herangezogen.

die unterrichtsbezogenen Lehrer- und Schüleraktivitäten fest, welche im Verlauf des Experimentierunterrichts auftraten. An dieser Stelle erhebt sich allerdings die Frage, ob es über die Definition der "Makro"- und "Mikrostruktur" hinaus notwendig ist, die unterrichtsbezogenen Aktivitäten des tatsächlichen Unterrichtsverlaufs zu protokollieren, zumal durch das Vorhandensein von Beobachtern die Reaktivität der experimentellen Situation erhöht und somit die externe Validität verringert wird. Die Protokollierung war einmal notwendig, um festzustellen, wie oft die einzelnen Schüler vom Lehrer direkt und unmittelbar verstärkt wurden und in welchem Ausmaß sie ausdrücklich auf Lehrerfragen bzw. Lehrerimpulse reagierten, da die Kenntnis dieser Sachverhalte zur Klärung unserer Hypothesen 15, 16, 17 und 18 notwendig war. Zum anderen sind wir der Ansicht, daß die Kenntnis der Verteilung eindeutig definierter unterrichtsbezogener Lehrer- und Schüleraktivitäten für eine differenzierte Interpretation der Versuchsergebnisse herangezogen werden kann. Weiter besteht im Fall einer totalen oder partiellen Wiederholung unserer Untersuchung die Möglichkeit, den tatsächlichen Unterrichtsablauf wenigstens bezüglich einiger definierter Kriterien zu vergleichen und gegebenenfalls abweichende Ergebnisse daraufhin zu interpretieren. Schließlich bietet uns die Protokollierung operational definierter Aspekte des tatsächlichen Unterrichtsverlaufs die Möglichkeit, die unter Punkt 4.31 geschilderten Bedingungen - jedenfalls teilweise - einer nachträglichen empirischen Kontrolle zu unterziehen. So ist es z.B. möglich, die unter Punkt 4.31 a) behauptete gleichmäßige Aktivitätsverteilung zwischen Lehrern und Schülern anhand der Unterrichtsprotokolle zu quantifizieren. Zusammenfassend meinen wir daher festhalten zu können,

daß ungeachtet des experimentellen Aufwands und der Verringerung der externen Validität die Protokollierung der unterrichtsbezogenen Lehrer- und Schüleraktivitäten gerechtfertigt ist.

7.71 Die Mikrostruktur unseres "herkömmlichen Vergleichsunterrichts"

Die Lehrziele des "herkömmlichen Vergleichsunterrichts" wurden mit den Lehrzielen der beiden Programme abgestimmt. Bei dieser Lehrzielabstimmung begnügten wir uns nicht damit, lediglich formale Unterrichtsanweisungen zu geben (wie z.B. "der Schüler muß Wurzeln addieren bzw. subtrahieren können"), vielmehr wurden die Lehrer veranlaßt, inhaltlich von den einzelnen Lernschritten des Programms auszugehen. Dadurch war gewährleistet, daß eine optimale inhaltliche Übereinstimmung der verglichenen Unterrichte bestand. Wir wollen im folgenden anhand eines Lernschrittes¹⁾ exemplarisch aufzeigen, in welchem Bereich der Lehrer freie Hand hatte und wo die Voraussetzungen dafür geschaffen worden sind, daß die verglichenen Unterrichtsmethoden bezüglich des Lehrinhaltes weitgehend übereinstimmen:

Die Information, daß Exponenten einer Potenz positive oder negative, ganze oder gebrochene Zahlen sein können und der Lehrsatz, daß Bruchpotenzen potenziert werden, indem man die Exponenten multipliziert, blieben bezüglich der Gestaltung völlig dem Lehrer überlassen. Die daran im Programm anschließende Übungsaufgabe zur An-

1) Es handelt sich dabei um den Lernschritt 2 der Lehrereinheit 2 im Programm "Rechnen mit Wurzeln".

wendung des Gelernten ("Wie groß ist dann $(16^{1/3})^3$?")
m u ß t e Bestandteil des herkömmlich geführten Un-
terrichts sein. Um diese Aufgabe lösen zu können, mußte
vom Lehrer "automatisch" die oben erwähnte Information
und der Lehrsatz geboten werden. Frei war er allerdings
bezüglich der Beispiele, mit welchen er diese Infor-
mation einführte und den Lehrsatz demonstrierte sowie
in der Erarbeitung der Information bzw. des Lehrsatzes,
d.h. also der logischen Verknüpfung der Beispiele und
des Ausmaßes der Schülerbeteiligung an der Erarbeitung
(so konnte der Lehrsatz z.B. informativ geboten werden,
wie dies auch im Programm geschieht, oder der Schüler
konnte dazu veranlaßt werden, ihn selbsttätig zu "ent-
decken"). Weiter hatte der Lehrer die Möglichkeit, die
logische Sequenz der Lernschritte des Programms und
der Lerneinheiten in seinem Sinne abzuändern¹⁾. Schließ-
lich hatte der Lehrer absolute Freiheit, falsche Schü-
lerantworten überhaupt nicht oder mit selbstgewählten
Beispielen zu korrigieren. Die Freiheit in der Wahl der
logischen Sequenz der einzelnen Lernschritte bzw. der
Abfolge übergeordneter Lernziele ist Voraussetzung für
die "Beweglichkeit" des "herkömmlichen Vergleichs-
unterrichts", d.h. also der Möglichkeit, während des Un-
terrichts "vor-" und "zurückzuspringen", sich an Schü-
lerantworten anzupassen usw. Während der Zeit der Un-
tersuchung wurde auf Hausaufgaben verzichtet, um einen
unkontrollierten und zusätzlichen Übungsgewinn zu ver-
meiden. Die Übungsfunktion der Hausaufgaben mußte also in
den Unterricht mit einbezogen werden. Schließlich wur-
de eine einheitliche Unterrichtsdauer von 5 Schulstun-
den festgelegt, welche ohne Rücksicht darauf, ob der

1) Von dieser Möglichkeit wurde durchweg Gebrauch ge-
macht.

Lehrer den Stoff umfangmäßig bewältigte oder nicht, eingehalten werden mußte. Die Prüfungsaufgaben waren dem Lehrer nicht bekannt. Angesichts der Tatsache, daß der Lernstoff im Rahmen der Übungsaufgaben und der damit notwendigerweise verbundenen Informationen streng und eindeutig definiert war¹⁾, konnten wir auf die Bekanntgabe der Prüfungsaufgaben verzichten und dadurch vermeiden, daß die Lehrer absichtlich oder unabsichtlich Aufgaben in ihren Unterricht einbauten, welche mit den Prüfungsaufgaben nicht nur strukturgleich sondern darüber hinaus inhaltlich identisch gewesen wären²⁾.

-
- 1) So waren z.B. im Unterricht "Rechnen mit Wurzeln" 67 Übungsaufgaben bei beiden Versuchsgruppen identisch. Daraus wird ersichtlich, wie weitgehend die Steuerung der inhaltlichen Übereinstimmung erfolgte, zumal gerade diesen Übungsaufgaben eine Schlüssel-funktion zukam. Daneben hatten die Lehrer in den konventionell unterrichteten Klassen selbstverständlich die Möglichkeit, noch eine Vielzahl anderer Übungsaufgaben selbständig zu wählen, wovon auch durchweg Gebrauch gemacht wurde. Den mit dem Programm arbeitenden Schülern standen als weitere Übungsaufgaben die am Ende jeder Programmeinheit aufgenommenen Wissensfragen zur Verfügung.
 - 2) Die Lehrer wurden zusätzlich zu dem oben Ausgeführten aufgefordert, jeden Schüler möglichst immer beim Namen aufzurufen, wenn sie Gelegenheit zu einer ausdrücklichen Reaktion geben. Diese Maßnahme war Voraussetzung für eine zuverlässige Protokollierung des Schülerverhaltens und der darauf bezogenen Lehrer-verstärkungen. Entsprechende Erfahrungen wurden während der Vorversuche gemacht. Ohne diese Vorkehrung wäre es nicht immer möglich gewesen, einen angesprochenen bzw. antwortenden Schüler zuverlässig zu lokalisieren, da das Antwortverhalten der Schüler häufig nur im Nennen einer Zahl bestand und er zudem dem dafür zuständigen Beobachter den Rücken zukehrte.

Wir meinen mit dieser Form der Definition der "Mikrostruktur" unseres "herkömmlichen Vergleichsunterrichts" auf der einen Seite die Voraussetzung dafür geschaffen zu haben, daß die Unterrichte der verglichenen Gruppen bezüglich des vermittelten Stoffes auch bei strengen Anforderungen vergleichbar sind, daneben aber dem "herkömmlichen Vergleichsunterricht" jenen freien Spielraum gewährt zu haben, der ihn ja als ein potentiell beweglicheres System von der Programmierten Unterweisung abhebt.

7.72 Protokollierung des Unterrichtsablaufs

a) Beobachter:

Für die Beobachtung und Protokollierung des "herkömmlichen Vergleichsunterrichts" der sechs Klassen standen 6 Beobachter zu Verfügung. Fünf Beobachter waren Lehrerstudenten aus dem 5. Semester, ein Beobachter war der Doktorand selbst. Sämtliche Beobachter waren aufgrund ihres bisherigen Studienganges bzw. der Tätigkeit an einer Pädagogischen Hochschule mit dem Ablauf eines konventionellen Unterrichtsgeschehens vertraut, d.h. sie hatten gelernt, sich auf die Aktionsformen von Lehrern und Schülern einzustellen.

b) Technische Hilfsmittel:

Der Ablauf des Unterrichtsgeschehens wurde zusätzlich zu der Beobachtung mit Hilfe eines Tonbands akustisch festgehalten. Die Tonbandaufnahmen ermöglichten es, das auf den Beobachtungsbogen festgehaltene Lehrer- und

Schülerverhalten teilweise noch einmal nachträglich zu überprüfen¹⁾. Die Beobachter saßen am rückwärtigen Ende des Klassenzimmers, so daß sie während des Unterrichts von den Schülern nicht gesehen werden konnten. Vor Ablauf des Hauptversuchs wurden die Schüler während einer zwei- bzw. dreistündigen Eingewöhnungsphase mit der Anwesenheit der Beobachter vertrautgemacht. Nach übereinstimmendem Urteil von Lehrern und Beobachtern konnte am Ende der Eingewöhnungsphase keine Störung des Unterrichtsgeschehens mehr festgestellt werden. Jeder Beobachter verfügte über einen großflächig angelegten "Klassenspiegel", auf welchem jeder Schüler durch eine rechteckige Zelle namentlich verzeichnet war. Für jede zu beobachtende Verhaltenskategorie wurde ein Buchstabensymbol festgelegt, bei den Lehrerverhaltenskategorien wurde zusätzlich alle 10 Minuten eine zeitliche Markierung vorgenommen, diese diente später zur nachträglichen Überprüfung der Protokollierung schwieriger Unterrichtspassagen²⁾ mittels der Tonbandaufnahmen.

-
- 1) Eine vollständige Rekonstruktion des unterrichtlichen Geschehens war nicht möglich, da die Lehrer - trotz Aufforderungen - die Schüler nicht immer mit Namen anredeten und anhand der Tonbandaufnahmen nachträglich daher nicht mehr festgestellt werden konnte, welcher Schüler in solch einem Fall angesprochen worden war.
 - 2) Insbesondere dozierend vorgebrachte Unterrichtspassagen wurden mit einem Fragezeichen im Protokoll gekennzeichnet und nachträglich auf die Anzahl gesprochener Sätze hin überprüft. Die zeitlichen Markierungen dienten dabei als Orientierungshilfen.

c) Einübung der Beobachter und Organisation der Beobachtung:

Die sechs Beobachter wurden in zwei Gruppen mit je drei Beobachtern eingeteilt. In jeder Gruppe protokollierte ein Beobachter die Tätigkeitsformen des Lehrers, einer die Anzahl der Verstärkungen und einer die Schüleraktivitäten. Jeder der beiden Gruppen waren drei Klassen zugeteilt, wo sie während des Hauptversuchs insgesamt 15 Unterrichtsstunden zu protokollieren hatten. Vor Durchführung des Hauptversuchs protokollierte jede Gruppe in jeder der ihr zugewiesenen Klassen je zwei Stunden. Für diese Vorübungen wurden bereits Mathematikstunden der im Hauptversuch tätigen Lehrer gewählt. Nach dieser getrennten Einübung protokollierten alle sechs Beobachter gemeinsam 4 Schulstunden. Die während der gemeinsamen Sitzungen erstellten Protokolle wurden anschließend auf Übereinstimmung überprüft. Während der Vorversuche fanden laufend Besprechungen statt, in denen aufgetauchte Schwierigkeiten diskutiert und die Definitionen der Beobachtungskategorien präzisiert wurden.

d) Definition der Beobachtungskategorien:

Die Beobachtungskategorien waren grundsätzlich so definiert, daß qualitative Unterschiede der Ausprägung einer Kategorie nicht erfaßt wurden. Dies wurde dadurch erreicht, daß sämtliche Beobachtungskategorien sich auf Verhalten bezogen, welches entweder vorhanden oder nicht vorhanden war und sich auf verbale Aktivitäten beschränkte. Als Maß für die Häufigkeit der Protokollierung einer Verhaltensform wurde die Satzeinheit gewählt, d.h. eine verbale Tätigkeitsform (z.B. eine Information) wurde dann

als einmal vorgekommen protokolliert, wenn ein in sich abgeschlossener Satz vorlag.¹⁾

Im einzelnen wurde folgendes protokolliert:

- 1) Unterrichtsbezogene Tätigkeitsformen des Lehrers (ohne Verstärkung):
 - 1.1) Informationen: Alle auf den behandelten Stoff bezogenen Aussagen, seien es nun Beispiele, Schlußfolgerungen oder Tätigkeiten wie Diktieren, Erklären, Wiederholen, Ergänzen, Zusammenfassen usw., sofern sie nicht im Rahmen der direkten Beantwortung einer Schülerfrage oder als Frage an die Schüler getätigt werden.
 - 1.2) Stoffliche Verlaufsäußerungen: Sämtliche verbalen Aktivitäten, welche sich darauf beziehen, die Schüler ausdrücklich zu motorischen oder intellektuellen Tätigkeiten zu aktivieren. Darunter fallen Äußerungen wie "schlagt eure Hefte auf", "paßt auf", "schaut an die Tafel" und "denkt nach" (intellektuelle "Tätigkeiten" im weitesten Sinn, darunter fällt also auch die Ausrichtung der Schülerkonzentration).

1) Ein durch den Lehrer bzw. einen Schüler unterbrochener Satz wurde als eine "einmal" vorgekommene Verhaltensform protokolliert. Schülerantworten bestanden meistens in einem mathematischen Ausdruck, häufig sogar nur in einer Zahl und wurden danach sofort wieder vom Lehrer unterbrochen. Unbesehen der Kürze der Aussage wurden sie trotzdem als eine "einmal" vorgekommene Verhaltensform protokolliert. Das Problem der Frage, wann ein Satz als abgeschlossen zu betrachten sei, stellte sich fast ausschließlich bei der Lehrerinformation, hier zeigten sich auch folgerichtig die größten Beobachterdiskrepanzen. Wir gingen davon aus, daß immer dann, wenn ein Punkt gesetzt werden könnte, eine Verhaltensform als "einmal" vorgekommen protokolliert werden sollte. D.h. zwei aneinandergereihte Hauptsätze wurden als zwei in sich abgeschlossene Sätze protokolliert.

- 1.3) Disziplinäre Verlaufsäußerungen: Alle verbalen Aktivitäten, welche "Ruhe und Ordnung" herzustellen beabsichtigen, also Äußerungen wie: "haltet den Mund" oder mehr indirekter Art wie "warum seid ihr heute so unruhig?".
 - 1.4) Fragen: Alle verbalen Aktivitäten, welche in Frageform an die Schüler gerichtet und stoffbezogen sind. Verbale Äußerungen, welche eine Frage näher erklären, fallen unter "Information".
 - 1.5) Antworten: Alle verbalen Aktivitäten, welche reaktiv in unmittelbarer zeitlicher Nachbarschaft auf eine Schülerfrage erfolgen und an den Fragesteller direkt gerichtet sind. Darunter fallen Bestätigung, Ablehnung und Informationen, sofern sie sich unmittelbar auf den Inhalt der gestellten Frage beziehen und wiederum ausschließlich an den Fragesteller gerichtet sind.
- 2) Unterrichtsbezogene Tätigkeitsformen der Schüler:
 - 2.1) Fragen und Antworten: Alle verbalen Schüleraktivitäten, welche an den Lehrer gerichtet sind. Erläuternde Bemerkungen zu den gestellten Fragen wurden hier ebenfalls protokolliert.
 - 2.2) Spontane Schülerinformation: Alle verbalen Schüleraktivitäten, die nicht reaktiv auf Lehrerfragen und Tätigkeitsaufforderungen seitens des Lehrers erfolgen. Negativ formuliert: jede verbale Schüleraktivität, welche stoffbezogen ist und nicht unter die Kategorie Frage bzw. Antwort fällt. Dazu rechnen spontane Korrekturen von Schüler- oder Lehreräußerungen, das Aufwerfen von Problemen usw.

3) **Verstärkungen:**

Unter Verstärkung wird jede verbale Aktivität seitens des Lehrers verstanden, welche eine stoffbezogene Schüleraktivität positiv bzw. negativ bewertet (Lob und Tadel) oder Auskunft darüber gibt, ob eine Schüleraktivität sachlich richtig oder falsch ist.

- 3.1) Positive Verstärkungen (individuell): Alle positiv bewertenden oder die sachliche Richtigkeit einer stoffbezogenen Schüleraktivität konstatierenden verbalen Äußerungen des Lehrers, die eindeutig auf einen Schüler gerichtet sind.
- 3.2) Negative Verstärkungen (individuell): Alle negativ bewertenden oder die sachliche Unrichtigkeit einer stoffbezogenen Schüleraktivität konstatierenden verbalen Äußerungen des Lehrers, welche eindeutig auf einen Schüler gerichtet sind.
- 3.3) Positive Verstärkungen (kollektiv): Alle unter die Definition von 3.1) fallenden verbalen Lehreraktivitäten, welche an mehr als einen Schüler gerichtet sind.
- 3.4) Negative Verstärkungen (kollektiv): Alle unter die Definition von 3.2) fallenden verbalen Lehreraktivitäten, welche an mehr als einen Schüler gerichtet sind.

7.73 Zuverlässigkeit der Erfassung und Validität der Beobachtungskategorien

a) Zuverlässigkeit der Erfassung

"Wer genaue Ergebnisse braucht, wird den Klassifizierungsvorgang so leicht machen, wie es der Zweck der Untersuchung erlaubt ... Bei jeder Untersuchung sollte zwischen dem Protokoll und dem Verhalten, das es repräsentieren soll, nur der einfachste Urteils- oder Unterscheidungsakt eingeschaltet werden: der, der nötig ist, um festzustellen, ob das Verhalten beobachtet werden konnte oder nicht"¹⁾. Wir haben diesen Weg eingeschlagen, indem wir von dem Kriterium "vorhanden" bzw. "nicht vorhanden" ausgingen²⁾, auf alle Aspekte des "mehr" oder "weniger Vorhandenseins" verzichteten und uns auf einwandfrei erkennbare verbale (im Gegensatz zu mimischen) Verhaltensformen beschränkten. Die Schwierigkeit der Protokollierung bestand kaum darin, eine Lehrer- bzw. Schüleraktivität zu klassifizieren als vielmehr darin, **a l l e** einschlägigen Aktivitäten zu erfassen. Hier ergab sich insbesondere die Schwierigkeit, bezüglich des für die Häufigkeit der Kategorien entscheidenden Kriteriums der "Satzeinheit" Übereinstimmung zwischen den Beobachtern herzustellen. Diese Schwierigkeit ergab sich allerdings ausschließlich für die-

1) W. SCHULZ, W.P. TESCHNER, J. VOIGT: "Verhalten im Unterricht. Seine Erfassung durch Beobachtungsverfahren", in: K.-H. INGENKAMP (Hrsg.): "Handbuch der Unterrichtsforschung Teil I", Weinheim-Berlin-Basel 1970, S. 655

2) Arbeiten mit entsprechenden Beobachtungskriterien werden von W. SCHULZ, W.P. TESCHER und J. VOIGT, a.a.O., S. 664 ff. berichtet.

jenigen Beobachter, welche das Lehrerverhalten zu protokollieren hatten¹⁾. Bei den Kategorien "Schüleraktivität" und "Verstärkungen" wurde die Protokollierung entscheidend dadurch erleichtert, daß die Lehrer die einzelnen Schüler mit Namen aufriefen. Trotz dieser von uns eingebauten Sicherung erforderte es auch hier seitens der Beobachter einige Geschicklichkeit, a l l e von ihnen zu protokollierenden Aktivitäten zu erfassen.

Die Zuverlässigkeit der Erfassung unserer Beobachtungskategorien wurde dadurch ermittelt, daß in der vorexperimentellen Phase in vier unserer sechs Versuchsklassen je eine Stunde diejenigen Beobachterpaare jeweils gemeinsam protokollierten, welche in der experimentellen Phase in verschiedenen Klassen denselben Sachverhalt zu protokollieren hatten.

Hierbei ergaben sich folgende Werte:

-
- 1) Tatsächlich ergab die Reliabilitätsprüfung, daß bei der Kategorie "Information" die geringste Beobachterübereinstimmung erzielt werden konnte. Es ist diejenige Kategorie, welche einmal am häufigsten protokolliert werden mußte und zum anderen längere verbale Sequenzen aufwies. An dieser Stelle tat sich für die Beobachter die Schwierigkeit auf, sich bezüglich der Frage, wann ein Satz zu Ende sei, zu einigen. Entsprechend groß sind die Diskrepanzen der protokollierten Häufigkeiten. Angesichts der Tatsache, daß die Lehrerverhaltenskategorien zur weiteren Verrechnung nicht herangezogen wurden sondern lediglich die oben bereits beschriebenen Funktionen erfüllte, ist die in diesem Bereich aufgefundene relativ geringe Zuverlässigkeit der Erfassung nicht gravierend.

	1. Std.	2. Std.	3. Std.	4. Std.
"Information"	81 %	72 %	75 %	84 %
"Stoffliche Ver- laufsäußerungen"	93 %	89 %	90 %	96 %
"Disziplinäre Ver- laufsäußerungen"	100 %	100 %	100 %	100 %
"Fragen"	90 %	92 %	94 %	91 %
"Antworten"	95 %	92 %	95 %	89 %

	Klasse 1 (31)	Klasse 2 (16)	Klasse 4 (22)	Klasse 6 (29)
"positive Ver- stärkungen" 1)	2	0	1	0
"Schülerakti- vitäten" 2)	3	1	0	0

Die Prozentwerte wurden aus dem Quotienten der von den beiden Beobachtern jeweils protokollierten Häufigkeiten gebildet, wobei die größere Zahl grundsätzlich in den Nenner kam. Dieser Quotient wurde mit 100 multipliziert. Die Prozentwerte bezeichnen also den Grad der Übereinstimmung in den protokollierten Häufigkeiten.

Die durchgehende absolute Übereinstimmung bei der Kategorie "disziplinäre Verlaufsäußerungen" erklärt sich daraus, daß in sämtlichen unterrichteten Klassen disziplinäre Verlaufssteuerungen äußerst selten waren,

-
- 1) Die Übereinstimmung bezüglich der negativen Verstärkungen wurde nicht errechnet, da diese bei der Formulierung der Hypothesen nicht weiter berücksichtigt wurden.
 - 2) Bei der Berechnung der Zuverlässigkeit der Erfassung der Schüleraktivitäten wurden die beiden Kategorien (vgl. 7.72) zusammengefaßt, da spontane Schüleraktivitäten nur vereinzelt vorkamen und bei der Prüfung der Hypothesen von sämtlichen protokollierten Schüleraktivitäten ausgegangen wurde.

selbst im ungünstigsten Fall entfielen pro Stunde nur 10 entsprechende Äußerungen. Zudem erfolgten diese in deutlich abgesetztem Tonfall und in knapper Form. In jedem Fall handelte es sich um eine mehr oder weniger markante Unterbrechung des Unterrichtsgesprächs. Die unterhalb der Klassen in Klammern gesetzten Zahlen bedeuten die Zahl der in der Beobachtungsstunde anwesenden Schüler. Die nicht in Klammer gesetzten Zahlen bedeuten die Anzahl derjenigen Schüler, bei denen die Häufigkeit der protokollierten Kategorien bei den beiden Beobachtern abweichen. Diese Abweichung betrug in der überwiegenden Anzahl der Fälle nur eine Verhaltenseinheit, nur in einem Fall wurde eine Differenz von 3 Verhaltenseinheiten festgestellt.

b) Validität

"Eine Prüfung der Inhaltsgültigkeit (Inhaltsvalidität) eines Beobachtungsverfahrens hätte zu sichern, ob wir damit messen, was wir messen wollen", es müßte also geprüft werden "... ob die Operationalisierung der Untersuchungsbegriffe wirklich zu Beobachtungen führt, die die in den Begriffen formulierten Untersuchungshypothesen zu bestätigen oder zu verwerfen imstande sind".¹⁾ Wir haben im Rahmen der Kapitel 4.11, 4.22 und 5.00 ausführlich die Bedeutung des Prinzips der Verstärkung und der aktiven ausdrücklichen Reaktion für Lernprozesse diskutiert und die entsprechenden Begriffe erläutert. Es braucht daher an dieser Stelle nicht mehr aufgewiesen zu werden, daß die Kategorien "Verstärkung" und "stoffbezogene Schüleraktivität" das messen, was wir messen wollen, zumal es sich hier um eindeutig definierte und direkt erfassbare Verhaltensweisen handelt. Es muß lediglich darauf hingewiesen werden, daß wir die nichtverbalen Verstärker wie zustimmendes Nicken,

1) W. SCHULZ, W.P. TESCHNER, J. VOIGT, a.a.O., S. 657

aufmunterndes Lächeln usw. nicht erfaßt und somit nicht a l l e Lehrerverhaltensweisen protokolliert haben, welche als direkte Verstärker wirksam werden können. Die Protokollierung von nichtverbalen Verstärkerreizen hätte jedoch einmal die Zuverlässigkeit der Erfassung nachhaltig beeinträchtigt, da diese nur selten eindeutig klassifizierbar und feststellbar sind, zum anderen sind solche verstärkenden Reize nicht nur für den Beobachter sondern auch für die Schüler nicht immer eindeutig interpretierbar. Das aber hat zur Folge, daß ihr tatsächliches Vorkommen kein genügend zuverlässiger Indikator für ihr Wirksamwerden ist. Schließlich kann noch darauf verwiesen werden, daß diese Verstärkerreize fast ausschließlich im Zusammenhang mit verbalen Verstärkerreizen auftreten, so daß ihnen primär die Funktion einer Modifizierung der Verstärkerqualität zuerkannt werden kann, ein Tatbestand, den - wie wir andernorts gesehen haben - SKINNER selbst für unwesentlich hielt, da gerade der menschliche Organismus selbst auf l e i c h t e (verstärkende) Hinweisreize reagiert. Bei der Kategorie "stoffbezogenes Schülerverhalten" haben wir infolge der Beschränkung auf verbal geäußertes Verhalten das schriftliche Ausarbeiten und Formulieren einer Lösung nicht protokolliert und somit ebenfalls auf eine "umfassende" Registrierung verzichtet. Da die Schüler aber fast keine schriftlichen Notizen machten und diese aufgrund der Tatsache, daß jeder Beobachter i.allg. über 20 Schüler gleichzeitig zu beobachten hatte, auch nicht annähernd zuverlässig hätten protokolliert werden können, haben wir auf deren Erfassung verzichtet. Zudem haben wir die entsprechenden Hypothesen so formuliert, daß sie sich ausdrücklich mit den auf sie bezogenen Beobachtungs-

kategorien decken, d.h. sie beziehen sich ausschließlich auf die Anzahl stoffbezogener verbaler Schüleräußerungen¹⁾. Bei der Formulierung der Lehrer-verhaltenskategorien gingen wir davon aus, daß mit wachsender einseitiger Information seitens des Lehrers die Möglichkeit aktiver Schülerbeteiligung in wachsendem Maße eingeschränkt wird. Der aktive Anteil der Schüler am Unterrichtsgeschehen wird weiter durch das Ausmaß der verbalen Aktivität des Lehrers im weitesten Sinne eingeschränkt. Wir haben daher zusätzlich zur Komponente der einseitigen Lehrerinformation die gesamte stoffbezogene verbale Lehreraktivität protokolliert. Da wir - wie wir unter Kapitel 5.00 geschildert haben, davon ausgehen, daß das Ausmaß an ausdrücklich verbaler Schüleraktivität nicht genügt, um das tatsächliche Ausmaß der "Mitarbeit" der Schüler zu erfassen, haben wir uns nicht darauf beschränkt, lediglich das unaufgeschlüsselte Verhältnis der Häufigkeit verbaler Lehrer- und Schüleraktivitäten zu erfassen, sondern haben darüber hinaus versucht, Lehrerverhaltenskategorien zu formulieren, welche auch einen indirekten Schluß auf das Ausmaß an "innerer", nicht direkt beobachtbarer Schüleraktivität zulassen. Wir gehen davon aus, daß mit wachsender Zahl stofflicher Verlaufsäußerungen die Unterrichtsbeteiligung der Schüler angeregt wird, wobei sich dies sowohl in innerer als auch in

1) Wir haben also keinesfalls beabsichtigt, die stoffbezogene verbale Schüleraktivität als eine "Eigenschaft" der betreffenden Schüler aufzufassen, welche sich in den ihr zugeordneten verbalen Reaktionen manifestiert. Eine solche "Eigenschaft" wäre nach unserer Auffassung eine hochkomplizierte intervenierende Variable, die man keinesfalls mit den von uns getroffenen Vorkehrungen erfassen könnte.

ausdrücklich geäußelter Aktivität manifestieren kann. Die stofflichen Verlaufsäußerungen sind nach unserem Dafürhalten ein Indikator für die Bereitschaft des Lehrers, die Klasse aktiv (und hierunter fällt auch der Begriff der "inneren Aktivität") am Unterrichtsgeschehen zu beteiligen. Dies gilt auch für die Häufigkeit von Lehrerfragen und die Beantwortung von Schülerfragen. Hier wird allerdings im Gegensatz zu den stofflichen Verlaufsäußerungen die Schüleraktivität direkt auf eine ungelöste Problemsituation bezogen. Die disziplinären Verlaufsäußerungen wurden als Beobachtungskategorie mit aufgenommen, um einen indirekten Hinweis darauf zu erhalten, in welchem Ausmaß der jeweilige Unterricht Störungen ausgesetzt war. Zusammenfassend muß ausdrücklich darauf hingewiesen werden, daß wir mit unseren Lehrerverhaltenskategorien in keiner Weise beabsichtigten, die Qualität unserer "herkömmlichen Unterrichte" zu erfassen, da dazu - allein was die Kognitionshilfen anbelangt - wenigstens 8 hochstrukturierte und kaum eindeutig klassifizierbare Kategorien des Lehrerverhaltens hätten erfaßt werden müssen.¹⁾ Wir können allerdings in Anspruch nehmen, das uns besonders interessierende Verhältnis der ausdrücklich geäußerten verbalen Aktivität von Lehrern und Schülern direkt und vollständig erfaßt zu haben.²⁾

- 1) Vgl. G. DIETRICH: "Die Auswirkungen von Unterrichtsstilen auf das Leistungs- und Sozialverhalten von Schülern", in: Blätter für Lehrerfortbildung. "Das Seminar", 2, 1971, S. 50
- 2) Die von uns aufgezeigte Inhaltsgültigkeit der Lehrerverhaltenskategorien kann indirekt durch die Arbeit von C.D. JAYNE: "A study of the relationship between teaching procedures and educational outcomes", in: Journ. exp. Educ., 1945, 14, S. 101-134, belegt werden. In dieser Studie wurde versucht, Skalen zu entwickeln, mit deren Hilfe der Zusammenhang von Lehrerverhalten und Unterrichtserfolg festgestellt werden kann. Der Autor kam dabei u.a. zu folgenden item-Gruppen: "Anzahl der Beiträge, die aus spontanem Schülergespräch erwachsen", "Anzahl der Erläuterungen durch den Lehrer", "Häufigkeit der sachlichen Information durch den Lehrer", "Häufigkeit von Lehrerfragen in bezug auf die Richtigkeit einer Schülerantwort" und "Häufigkeit, mit der eine Schülerantwort vom Lehrer als richtig bezeichnet wird". Die hier aufgezählten item-Gruppen weisen deutlich in die Richtung der Definition unserer Beobachtungskategorien. Weiter verweisen wir auf die Arbeit von W. SCHULZ, W.P. TESCHNER und J. VOIGT, a.a.O., S. 664 ff. Im Rahmen dieser Arbeit werden die wichtigsten einschlägigen amerikanischen Veröffentlichungen berichtet. Eine nähere Analyse zeigt, daß insbesondere die Form unserer Definitionen aber auch deren Inhalt mit den dort berichteten Ansätzen weitgehend übereinstimmen. Insbesondere wird deutlich, daß auf die Erfassung komplexer Verhaltensweisen mittels Ratingskalen zugunsten der einfachen Protokollierung von Häufigkeiten eindeutig definierten Lehrer- und Schülerverhaltens verzichtet wird.

7.80 Maßkriterien der Untersuchung

7.81 Unterrichtsbezogene Maßkriterien

7.811 Grundsätzliche Bemerkungen:

Aufgabe des Testinstrumentariums ist es, die Leistungsveränderungen, welche durch die planmäßige Variation der beiden Lehrmethoden bewirkt werden, zuverlässig und gültig zu messen. Im Gegensatz zu den sog. "normbezogenen" Tests, welche den Grad der Fähigkeit oder Fertigkeit, über welchen ein Individuum verfügt, "dadurch bestimmen, daß sie den Vergleich zu den Fähigkeiten oder Fertigkeiten anderer Individuen ermöglichen"¹⁾, handelt es sich in unserem Fall um einen sog. "kriteriumsbezogenen" Test²⁾. "Bezugspunkt für den Kriteriumstest ist ein 'absoluter Qualitätsstandard', die erwünschte Leistung oder das Endverhalten, das als Ziel oder Kriterium eines Lehrgangs, vor allem eines Lehrprogrammes, definiert ist"³⁾. Der absolute Qualitätsstandard wird dabei durch die v o l l - s t ä n d i g e n Lernziele des Programms repräsentiert.

-
- 1) K.-H. INGENKAMP: "Das Testen kognitiver Fähigkeiten und Leistungen", in: K.-H. INGENKAMP (Hrsg.): "Handbuch der Unterrichtsforschung, Teil I", Weinheim-Berlin-Basel 1970, S. 1025
 - 2) "Ein Test zur Erfassung der Effektivität eines Lehrprogramms ist ein 'kriteriumsbezogener' Test, insofern er auf einen absoluten Qualitätsstandard, die definierten Lernziele des Lehrprogramms, bezogen ist. Durch eine kriteriumsbezogene Leistungsmessung soll das tatsächliche Leistungsverhalten eines Schülers in Relation zum gewünschten Leistungsverhalten festgestellt werden können" (P. STRITTMATTER: "Präsentationsmodi von Lehrprogrammen", Weinheim 1970, S. 158 f.
 - 3) K.-H. INGENKAMP, a.a.O., S. 1025

Es ist daher bei der Testkonstruktion zuallererst darauf zu achten, daß alle vom Programm angezielten Lernziele möglichst differenziert im Test vertreten sind. Mit Hilfe solcher Kriteriumstests soll also festgestellt werden, ob und in welchem Ausmaß eine Schülergruppe, welche mit Hilfe eines Lernprogramms arbeitet, die Merkmale eines gewünschten Endverhaltens vollständiger erreicht als wenn der Lernstoff frontalunterrichtlich dargeboten wird. Der Rückgriff auf einen kriteriumsbezogenen Test hat nun für die Testkonstruktion Konsequenzen, die näher erläutert werden müssen. Bei den normbezogenen Tests wird meistens auf eine Teilpunktzahl verzichtet, da die aus ökonomischen Gründen relativ kurzen Testteile oft für sich allein keine genügende Reliabilität besitzen. Da jedoch beim Vergleich von Schülergruppen ein wesentlich geringerer Grad an Zuverlässigkeit in Kauf genommen werden kann¹⁾, haben wir unsere Gesamtzahl an items in die zwei Gruppen "knowledge" und "comprehension"²⁾ aufgeteilt und haben dadurch zwei "absolu-

1) Vgl. K.-H. INGENKAMP, a.a.O., S. 1024

2) In einem item zur Erfassung der Lernzielkategorie "knowledge" wird der Schüler aufgefordert, die gelernten Rechenregeln an einfachen Aufgaben anzuwenden. Um zur Lösung zu gelangen braucht er sich nur an die entsprechenden Beispiele im Programm bzw. im Lehrerunterricht wieder zu erinnern. Er muß also an einer vorliegenden Prüfungsaufgabe eine im Unterricht entsprechend eingeübte Aufgabe wiedererkennen. Prüfungsaufgabe und im Unterricht verwendete entsprechende Übungsaufgaben sind bezüglich ihres formalen Aufbaus identisch. Demgegenüber erfordert die Lernzielkategorie "comprehension" die Fähigkeit des Schülers, die an einfachen Übungsaufgaben demonstrierten Rechenregeln durch Verstehen auf items zu übertragen, welche kein Pendant im Lehrprogramm bzw. im Lehrerunterricht aufweisen. Es handelt sich hier also um einen Aspekt der Transferleistung. Die beiden Begriffe "knowledge" und "comprehension" gehen auf BLOOM zurück und werden in der neueren deutschen Literatur u.a. von STRITTMATTER (vgl. P. STRITTMATTER, a.a.O., S. 102 ff.) übernommen, wobei sich z.T. beträchtliche - gemessen an der BLOOMschen Definition - Begriffsverschiebungen ergaben.

te Qualitätsstandards" gewonnen. Weiter haben wir innerhalb unserer beiden Testteile die einzelnen items abgestuft bewertet, was zweifellos die Reliabilität der beiden Testteile einschränkt, da die Abstufungen nicht in dem Maße objektiv festgehalten werden können, wie dies bei einer "richtig - falsch" Bewertung der Fall wäre. Der von uns gewählte Weg schien uns trotzdem besser zu sein, weil es eine ungebührliche Vereinfachung wäre, sich auf die Feststellung zu beschränken, daß ein Schüler ein Unterrichtsziel erreicht oder nicht erreicht hat. Ein kriteriumsbezogener Test muß auch die Kategorie "teilweise erreicht" umfassen¹⁾. Wenngleich nun die Frage, inwieweit ein Teilschritt auf einem richtigen Lösungsansatz beruht, nicht immer eindeutig zu klären ist, haben wir das Prinzip der Gewichtung der Lösung jedes items zwischen den Punktwerten 0,00, 0,25, 0,50, 0,75 und 1,00 - abgesehen von der Prüfung des Vorwissens - durchgängig gewählt. Eine solchermaßen entstandene Reliabilitätsreduzierung muß nach unserer Auffassung bei kriteriumsbezogenen Tests sogar in Kauf genommen werden. Schließlich soll die Eigentümlichkeit eines kriteriumsbezogenen Tests noch an dem Problem der Schwierigkeitsindices der items aufgezeigt werden. In einem normbezogenen Test wird i.allg. auf Schwierigkeitsindices, welche über 0,90 liegen, verzichtet, weil solche Aufgaben zu leicht sind und daher zu wenig Trennschärfe aufweisen. Bei einem kriteriumsbezogenen Test wird z.T. geradezu davon ausgegangen,

1) Wir möchten dies an einem Beispiel kurz erläutern. Die Aufgabe: "Bestimme den Wert durch Subtraktion der Einzellogarithmen: $\log_4 (64:4)$ " hat die optimale Lösung $3 - 1 = 0$. Es wäre nun unangemessen, diese Lösung mit Punktwert 1,00 und die notwendige Zwischenlösung $\log_4 64 - \log_4 4$ mit Punktwert 0,00 zu gewichten. Der Schüler hat nämlich im letzten Fall einen sehr wesentlichen Schritt zur Aufgabenlösung bereits vollzogen.

daß 90 % der Schüler oder mehr am Ende des Lernprozesses das erwünschte Verhalten zeigen¹⁾. Entsprechend ist unser Test "knowledge" ausgerichtet: er soll ermitteln, inwieweit unsere mit verschiedenen Methoden behandelten Versuchsgruppen das gewünschte Endverhalten, nämlich das Beherrschen der einfachen Rechenfertigkeit mit Wurzeln bzw. Logarithmen aufweisen. Wenn gleich wir nicht von der Erwartung ausgegangen sind, daß 90 % der Schüler sämtliche Lernziele der Lernzielkategorie "knowledge" erreichen²⁾, haben wir doch bei der Auswahl der items primär darauf geachtet, ob sie einen Aspekt der oben angeführten einfachen Rechenfertigkeit repräsentieren, ungeachtet der Möglichkeit, daß solche items Schwierigkeitsindices haben können, welche über 0.90 liegen. Zusammenfassend glauben wir im Anschluß an INGENKAMP festhalten zu können, daß "bei der Entwicklung von Kriteriumstests ... die bisherigen Methoden der Aufgabenanalyse nur von sehr eingeschränktem Wert"³⁾ sind. Wir glauben daher auch mit gutem Grund auf

1) Vgl. K.-H. INGENKAMP, a.a.O., S. 1025 f.

2) Einmal ist dies bei deutschsprachigen Lehrprogrammen unseres Wissens nie der Fall, zum anderen haben wir, wenngleich der Trennschärfe nicht diejenige Bedeutung zukommen kann, welche sie im Rahmen von normbezogenen Tests hat, durch Übernahme von items aus dem Bereich der Schwierigkeitsindices 0,90 - 0,50 darauf geachtet, daß auch in unserem kriteriumsbezogenen Test soweit als möglich optimale Voraussetzungen für die Trennschärfe und damit die Reliabilität geschaffen werden. Auch ein kriteriumsbezogener Test kann es sich nicht leisten, auf solche Kriterien völlig zu verzichten, wenngleich man bei der Testkonstruktion aus eigengesetzlichen Gründen häufig gezwungen ist, diese Aspekte zu vernachlässigen. Ein entsprechendes Vorgehen findet sich u.a. bei STRITTMATTER (vgl. P. STRITTMATTER, a.a.O., S. 162 ff.)

3) K.-H. INGENKAMP, a.a.O., S. 1025

die Berechnung von Trennschärfeindizes und Reliabilitätskoeffizienten verzichten zu können, da eine solche Information nur von sehr beschränktem Wert wäre und wir, wie oben ausgeführt wurde, z.T. bewußt auf die Voraussetzungen für hohe Trennschärfeindizes und optimale Reliabilität verzichtet haben.

7.812 Test zur Ermittlung der Vorkenntnisse:

Dieser Test besteht aus 18 items, zu deren Bearbeitung 45 Minuten zur Verfügung standen. Er diente zur Ermittlung des Wissensbestandes im Potenzrechnen, da dieser eine bedeutsame Variable für das Eingangsverhalten unserer Versuchspersonen darstellt, denn Radizieren und Logarithmieren sind beide Umkehrungen des Potenzierens¹⁾. Weiter wurde mit Hilfe dieses Tests überprüft, ob und in welchem Ausmaß Schüler unserer Versuchsgruppen bereits über direkte Kenntnisse im Radizieren bzw. Logarithmieren verfügten. Die dazu verwendeten items sind so formuliert, daß selbst Schüler, die sich außerhalb der Schule oder durch zufällige Kontakte mit Schülern aus oberen Klassen einschlägige bruchstückhafte Informationen verschafft hatten, diese richtig beantworten konnten. Da wir Schüler, die auch nur ein item der Nummer 9 - 14 gelöst hatten,²⁾ zwar im Versuchsunterricht mitlaufen ließen, in die endgültige Stichprobe jedoch nicht

-
- 1) Potenzieren: $a^n = x$; Lösung: $x = a^n$
Radizieren: $x^n = b$; Lösung: $x = \sqrt[n]{b}$
Logarithmieren: $a^x = b$; Lösung: $x = \log_a b$

(Vgl. H. FIEBIG und H. JUNGHANS: "Rechnen mit Logarithmen", o.J., S. 11)

- 2) Vgl. S. 248

aufnahmen¹⁾), achteten wir nicht weiter darauf, in welchem **A u s m a ß** direkte Kenntnisse des Wurzelrechnens bzw. des Rechnens mit Logarithmen vorhanden waren. Uns interessierte lediglich, ob **ü b e r h a u p t** einschlägige Kenntnisse vorlagen. Aufgrund unserer endgültigen Stichprobenauswahl konnten wir daher auch darauf verzichten, den Schlußtest gleichzeitig zur Überprüfung des Eingangsverhaltens heranzuziehen und dann als Maß die Lernzuwachsrate zu benutzen. Dies schien uns umso sinnvoller zu sein, als die Verwendung der Lernzuwachsrate mit einigen Schwierigkeiten behaftet ist²⁾. Die Vorkenntnisse im Potenzrechnen wurden in folgenden Bereichen überprüft: Addition und Subtraktion von Potenzen, Multiplikation und Division von Potenzen mit gleicher Basis, Multiplikation und Division von Potenzen mit gleichen Exponenten und Potenzieren von Potenzen. Punktwert für eine richtige Lösung war 1,0; eine Zwischengewichtung wurde hier nicht vorgenommen. Halbrichtige Lösungen wurden als falsch bewertet. Da der Vortest lediglich dazu diente, Schüler mit direkten Vorkenntnissen der im Versuch behandelten Stoffgebiete und Schüler, deren Leistungen im Potenzrechnen extrem von den Maßzahlverteilungen unserer Gruppen abwichen, aus der verrechneten Stichprobe auszuschneiden, sowie einen Hinweis darüber geben sollte, inwieweit unsere Gruppen A und B bezüg-

-
- 1) Wir haben mit diesem Vorgehen das von KULKARNI geschilderte Konstruktionsprinzip optimal erfüllt, demzufolge in einen kriteriumsbezogenen Test nur solche items aufgenommen werden sollen, welche von der noch ungeübten Gruppe nicht gelöst werden können (vgl. S.S. KULKARNI: "Kriteriumstests und Programmiertes Lernen", in: K.-H. INGENKAMP und Th. MARSOLEK (Hrsg.): "Möglichkeiten und Grenzen der Testanwendung in der Schule", Weinheim-Berlin-Basel, 1968).
 - 2) Vgl. K.J. KLAUER: "Über die Erfassung der Validität von Lernprogrammen", in: pl 1, 1968, S. 3

lich indirekter Vorkenntnisse vergleichbar sind, schien uns dieses vereinfachte Vorgehen ausreichend zu sein. Dies gilt umso mehr, als aufgrund unseres Kreuzvalidierungskonzepts jede Gruppe bezüglich ihres Lernerfolgs in den zwei Lehrmethoden mit sich selbst verglichen wurde und somit auch bezüglich indirekter Vorkenntnisse bei dieser "Vergleichsrichtung" von absolut vergleichbaren Bedingungen ausgegangen werden kann. Würden allerdings, wie dies i.allg. der Fall ist, lediglich zwei unabhängige Stichproben miteinander verglichen, müßten die Vorkenntnisse differenzierter geprüft werden. Hier wäre ein solch schematisches Auswertungsverfahren unangebracht¹⁾. Pro item standen durchschnittlich 2,5 Minuten Bearbeitungszeit zur Verfügung²⁾. Angesichts der großen Zahl an items, welche einen relativ geringen Schwierigkeitsindex haben und somit mit einem deutlich geringeren Zeitaufwand gelöst werden konnten, kann von einem Niveautest mit Speedkomponente ausgegangen werden. Da bei den üblichen Schulprüfungen i.allg. eine Speedkomponente enthalten ist, haben wir uns bei der Konstruktion unserer Tests auch daran gehalten. Ein Lernziel ist nämlich nicht nur dadurch gekennzeichnet, daß ein Schüler eine bestimmte Aufgabe bewältigen muß, sondern auch dadurch, daß dies in einem definierten Zeitraum geschieht.

-
- 1) Uns ist allerdings keine Vergleichsstudie bekannt, welche versucht hätte, i n d i r e k t wirksames Eingangsverhalten zu erfassen. Dieser Verzicht ist umso bedenklicher, als meist von einer massiven Stichprobenverzerrung aufgrund des Rückgriffes auf natürliche Klasseneinheiten ausgegangen werden muß.
 - 2) Der für die einzelnen items verwendete Zeitaufwand war den Schülern freigestellt.

7.813 Das Problem der inhaltlichen Formulierung der Testitems:

Bei der inhaltlichen Gestaltung der Testitems wurde sorgfältig darauf geachtet, keine items zu verwenden, welche im Programm auftauchen. Dies hätte eine ungerechtfertigte Bevorzugung der Programmgruppe gegenüber der herkömmlich unterrichteten Gruppe zur Folge gehabt und hätte somit zu einer Scheinüberlegenheit der Programmgruppe geführt, wie man sie bei vergleichbaren Untersuchungen häufig antreffen kann¹⁾. Die items wurden vielmehr von zwei an der Untersuchung nicht beteiligten Mathematiklehrern und dem Doktoranden eigens für die Prüfung erstellt. Ausgangspunkt waren dabei allerdings die unter Punkt 7.60 berichteten Lernziele der verwendeten Unterrichtsprogramme, d.h. die items wurden im Hinblick darauf entworfen, ob sie auf der Basis des im entsprechenden Programm dargebotenen Lernstoffs gelöst werden können. Entsprechend der Fragestellung der Untersuchung wurden dabei items für die Lernzielkategorien "knowledge" und "comprehension" erstellt. Kriterien für die Zugehörigkeit der items zu einer der beiden Lernzielkategorien waren, abgesehen von dem jedem einzelnen item zugrunde liegenden Lernziel, ihre Schwierigkeitseinstufung durch die beiden Mathematiklehrer, welche bezüglich der Unterrichtung dieses Stoffgebiets in unserer Stichprobe vergleichbaren Klassen über eine langjährige Berufserfahrung verfügten und der Zeitaufwand, welcher zur Berechnung der einzelnen items notwendig war. Zur Ermittlung des Zeitaufwands wurden die Beobachter der Untersuchung, welche vor Beginn des

1) Vgl. G. HUBER: "Lernpsychologische Befunde bei Programmierter Unterweisung", München 1967

Hauptversuchs die beiden Programme durcharbeiten mußten¹⁾, aufgefordert, die von den Mathematiklehrern und dem Doktoranden formulierten items zu berechnen und die für die einzelnen items notwendige Zeit zu notieren. Aus einer Liste von ursprünglich 150 items wurden schließlich 94 ausgewählt. Diese wurden in die Gruppen "Wurzelrechnen" und "Rechnen mit Logarithmen" aufgeteilt und nach Beendigung der jeweiligen experimentellen Phase unseren Versuchsklassen als Abschlußtest vorgelegt. Die den beiden Lernzielkategorien zugehörigen items wurden den Schülern nicht getrennt sondern gemischt vorgelegt. Mit Hilfe der empirisch ermittelten Schwierigkeitsindices der 94 items wurden schließlich diejenigen 70 items ausgewählt, welche zur endgültigen Verrechnung mit herangezogen wurden. Zusammengefaßt kann der Weg der inhaltlichen Auswahl der zur Überprüfung unserer Hypothesen herangezogenen items wie folgt dargestellt werden:

- a) Erstellung einer Ausgangsliste von 150 items nach den Lernzielen der Untersuchungsprogramme.
- b) Feststellung des zur Berechnung jedes einzelnen items notwendigen Zeitaufwands und qualitative Analyse der aufgrund der studentischen Berechnung aufgetauchten Schwierigkeiten.
- c) Zusammenstellung der in unseren Versuchsklassen verwendeten Anschlußtests (insgesamt 94 items).

1) Die Durcharbeitung der in der Untersuchung verwendeten Programme durch die Beobachter schien uns notwendig zu sein, um zu gewährleisten, daß die Beobachtung nicht durch den Beobachtern unbekannte Stoffe beeinträchtigt wird. Eine nähere Analyse der Beobachtungskategorien zeigt nämlich, daß eine mangelhafte Beherrschung des behandelten Stoffes an verschiedenen Stellen zu einer Beeinträchtigung der Zuverlässigkeit der Beobachtung führen muß.

- d) Ermittlung der Schwierigkeitsindices der 94 items an unserer Stichprobe¹⁾ und Zusammenstellung von 70 der der Überprüfung unserer Hypothesen zugrunde gelegten items nach dem Kriterium vergleichbarer mittlerer Schwierigkeitsindices (vgl. 7.815).

Die der Lernzielkategorie "knowledge" zugehörigen items entsprechen bezüglich ihres Komplexitätsgrades den im Programm verwendeten Rechenbeispielen, d.h. alle zu ihrer Lösung notwendigen Voraussetzungen waren im Verlauf des Unterrichts eingeübt worden. Die der Lernzielkategorie "comprehension" zugehörigen items unterschieden sich von den items der Lernzielkategorie "knowledge" dadurch, daß sie

- im herkömmlichen Unterricht bzw. im Programm nicht direkt (was den f o r m a l e n Aufbau anbelangt) behandelt wurden;
- mehr Zeitaufwand zur Bearbeitung bedurften;
- das "Verständnis" des gelernten Stoffes zur Lösung voraussetzten, d.h. daß eine rein mechanische Übertragung der im Versuchsunterricht eingeübten Rechenoperationen nicht möglich war;
- häufig die Beachtung mehrerer der gelernten Rechenoperationen innerhalb des Lösungsweges erforderten;
- oft "Fallen" enthielten, welche bei oberflächlichem Verständnis zu Fehlern führen mußten.

In jedem Fall genügte zur Lösung der items der Lernzielkategorie "comprehension" nicht die einfache Übertragung gelernter Rechentechniken. Jedes der items dieser Lernzielkategorie wies mindestens eine, meistens mehrere der oben genannten Bedingungen auf.

1) Die Ermittlung der Schwierigkeitsindices an unserer Versuchsstichprobe war unumgänglich, da nur so die Gewähr der direkten Vergleichbarkeit der Maßzahlen ein und derselben Gruppe geboten war (vgl. 7.30 und 7.815).

7.814 Gewichtung der items:

Die Gewichtung der items erfolgte so, daß jedes item u n a b h ä n g i g von seinem Schwierigkeitsindex mit dem Höchstwert 1,0 bewertet wurde. Auf die Alternativgewichtung "richtig - falsch" wurde aus bereits oben erwähnten Gründen verzichtet. Die Abstufungen der Bewertung lagen bei 1,0; 0,75; 0,50; 0,25 und 0,00. Die Bewertung 1,0 wurde nur gegeben, wenn die Aufgabe bis zur einfachst möglichen Form durchgerechnet wurde¹⁾. Der Punktwert 0,75 wurde gegeben, wenn der eigentliche Lösungsprozeß zwar abgeschlossen war, eine mathematisch "einfachere" Form jedoch erreichbar gewesen wäre (z.B. durch "Kürzen"). Der Wert 0,50 wurde gegeben, wenn ein Schüler zu einer Teillösung gelangt war, dann aber falsch weiterrechnete. Der Punktwert 0,25 wurde gegeben, wenn ein richtiger Lösungsansatz erkennbar war, dem Schüler aber bereits im Anfangsstadium der Aufgabenbearbeitung Fehler unterliefen²⁾. Die hier angebotene Beschreibung der Gewichtungskriterien muß sich auf formale Aspekte der Aufgabenlösung beschränken, sie soll daher unten

-
- 1) Eine solche lag vor, wenn das Lernziel erreicht und weiteres Vereinfachen durch "Kürzen" nicht mehr möglich war.
 - 2) Die Gewichtung der einzelnen items erfolgte durch die beiden an der Untersuchung nicht beteiligten Mathematiklehrer sowie den Doktoranden. Die Korrektur der Schülerarbeiten, an welcher die beiden Mathematiklehrer nicht mehr beteiligt waren, erforderte allerdings sehr häufig zusätzliche Entscheidungen, da es auch nicht annähernd möglich war, alle empirisch vorgekommenen Lösungsansätze und Teillösungen vorauszu-
sehen.

an einem Beispiel verdeutlicht werden¹⁾. Gleichzeitig muß darauf hingewiesen werden, daß es auch nicht annähernd möglich ist, sämtliche denkbaren Teillösungen aufzuführen, da es eine große Anzahl entsprechender Ansätze selbst für ein einziges Item gibt. Bei der Vielzahl der möglichen Teillösungen gab es vereinzelt Lösungsansätze, bei denen es im Ermessen des Auswertenden lag, welche Gewichtung zu wählen war. Um trotzdem Willkürentscheidungen des Auswertenden zu vermeiden, wurden sämtliche Arbeiten unabhängig vom Doktoranden von einer wissenschaftlichen Hilfskraft²⁾ korrigiert. Im Falle abweichender Gewichtungsvorschläge wurde der fragliche Lösungsansatz sorgfältig analysiert und anschließend eine gemeinsame Entscheidung getroffen. Bei Nichtübereinstimmung trotz erfolgter qualitativer Analyse wurde grundsätzlich die niederere Gewichtung gewählt.

-
- 1) Der mathematische Ausdruck $\frac{p\sqrt{q} + q\sqrt{p}}{\sqrt{pq}}$ soll so umgeformt werden, daß im Nenner die Zahl 1 steht. Wenn der Schüler bemerkt, daß der vorliegende Bruch mit dem Faktor $\sqrt{pq}$ multipliziert werden muß, die Multiplikation aber fehlerhaft erfolgt, erhält er 0,25 Punkte. Gelangt er zu dem Ausdruck $\frac{pq\sqrt{p} + pq\sqrt{q}}{pq}$, erhält er 0,50 Punkte. Klammert er den Ausdruck pq im Zähler aus und gelangt so zu dem Ausdruck $\frac{pq(\sqrt{p} + \sqrt{q})}{pq}$, erhält er 0,75 Punkte. Der Ausdruck $\sqrt{p} + \sqrt{q}$ wird mit 1,0 Punkten bewertet.
- 2) Diese hatte vor Beginn der Korrektur die Untersuchungsprogramme durchgearbeitet und zudem in mehreren Arbeitssitzungen mit dem Doktoranden den Lösungsschlüssel diskutiert.

7.815 Herstellung der Testendform:

Entsprechend unserem Versuchsplan stellten wir aus den endgültig ausgewählten 70 items vier Tests zusammen, welche der Berechnung unserer Hypothesen zugrunde liegen.

Die Tests weisen folgende statistische Kennwerte auf:

Radizieren	Logarithmieren
Lernzielkategorie "knowledge"	Lernzielkategorie "knowledge"
20 items mittl.P: 0,7225 s_p : 0,109 mittl.Punktzahl: 14,515	20 items mittl.P: 0,7220 s_p : 0,089 mittl.Punktzahl: 14,375
Lernzielkategorie "comprehension"	Lernzielkategorie "comprehension"
15 items mittl.P: 0,2500 s_p : 0,110 mittl.Punktzahl: 3,8007	15 items mittl.P: 0,2473 s_p : 0,140 mittl.Punktzahl: 3,6466

Dabei bedeuten:

mittl.P = arithmetisches Mittel der Schwierigkeitsindices

s_p = Streuung der Schwierigkeitsindices

mittl.Punktzahl = arithmetisches Mittel der erreichten Punktzahl (Rohwerte)

Die Schwierigkeitsindices der vier Tests wurden auf ihre Güte der Anpassung an die Normalverteilung überprüft. Für die je 20 items umfassenden Tests zur Erfassung der Lernzielkategorie "knowledge" wurde die Güte der Anpassung mittels des Chi-Quadrat-Tests überprüft. Bei den Tests zur Erfassung der Lernzielkategorie "comprehension" wurde der kombinatorische Test zur Güte der Anpassung herangezogen¹⁾. Die Nullhypothese konnte grundsätzlich nicht verworfen werden,

-
- 1) Dazu wurde die Gesamtfläche der beobachteten Verteilung entsprechenden theoretischen Verteilung in fünf flächengleiche Anteile zerlegt. Die tatsächlich beobachtete und die theoretisch erwartete Häufigkeit innerhalb der fünf flächengleichen Anteile wurden der Berechnung der Prüfgröße Chi-Quadrat zugrunde gelegt (vgl. G.A. LIENERT: "Verteilungsfreie Methoden in der Biostatistik", Meisenheim 1962, S. 74 ff.). Bei den Tests zur Erfassung der Lernzielkategorie "comprehension" konnte dieses Verfahren nicht mehr angewandt werden, da nur Erwartungswerte von eben noch 5 pro Flächenanteil bei der Anwendung der Chi-Quadrat-Methode noch zulässig sind. Bei den 15 items umfassenden Tests hätte dies nur die Bildung von 3 flächengleichen Teilen ermöglicht und somit zu einem sehr ungenauen Verfahren geführt. Die Güte der Anpassung an die Normalverteilung wurde daher mit dem "kombinatorischen Test" für die Güte der Anpassung überprüft. Dazu wurde die Kurvenfläche in 15 gleiche Teile zerlegt und der Erwartungswert 1 pro Flächenanteil mit dem tatsächlich beobachteten Wert verglichen. Die Nullklassen wurden mit den besetzten Klassen ins Verhältnis gesetzt (vgl. G.A. LIENERT, a.a.O., S. 76 ff.).

wir können daher bezüglich der Schwierigkeitsindices von einer normalverteilten Grundgesamtheit ausgehen¹⁾. Aus der Tabelle folgt weiter, daß die Tests der Lernzielkategorie "knowledge" bzw. "comprehension" bezüglich ihres mittleren Schwierigkeitsgrades vergleichbar sind. Da wir die Lernleistung einer Gruppe bei Zugrundelegung verschiedener Unterrichtsmethoden vergleichen wollen (vgl. 7.31), waren wir, um unnötige Maßzahltransformationen und Korrekturgrößen zu vermeiden und mit einer eindeutig interpretierbaren Maßzahl-differenz ($\bar{X}_{\text{Programm}} - \bar{X}_{\text{konventionell}}$) arbeiten zu können, auf die Konstruktion von Tests angewiesen, deren Rohwerte direkt verglichen werden können²⁾. Zu diesem

-
- 1) Bei den Tests zur Erfassung der Lernzielkategorie "knowledge" ergab sich beide Male ein Chi-Quadrat-Wert von 4,00. Bei den Tests zur Erfassung der Lernzielkategorie "comprehension" ergaben sich beide Male P-Werte von 0,2784. Die detaillierten Flächen-einteilungen, welche diesen Werten zugrunde liegen, finden sich im Anhang.
 - 2) Es wäre zwar grundsätzlich möglich gewesen, mit Hilfe von mehr oder weniger komplizierten Maßzahltransformationen ein $\bar{X}$ (korrigiert) zu konstruieren, welches auch den Vergleich von Leistungen bei Tests erlaubt hätte, welche bezüglich des mittleren Schwierigkeitsgrads - und somit (bei gleicher Gewichtung) gleichzeitig des arithmetischen Mittels der Rohwerte - voneinander abweichen. Solchermaßen korrigierte Maßzahlen sind allerdings i.allg. von geringerer Aussagekraft als unkorrigierte Maßzahlen, in jedem Fall sind es jedoch psychologisch schwer interpretierbare Größen, hinter denen sich im negativen Fall sogar statistische Artefakte verbergen können. Aus diesem Grund haben wir in der Anlage der Untersuchung streng auf direkt vergleichbare Maßzahlen geachtet. Angesichts der bestehenden Sachlage bot sich hier der von uns gewählte Weg an. Eine - durchaus noch eindeutig interpretierbare - Transformation in Standardwerte war nicht möglich, da wir aufgrund unseres experimentellen Aufbaus nicht davon ausgehen konnten, daß die mittels eines Standardwertes aufeinander bezogenen Maßzahlen einer Population angehören.

Zweck boten wir - wie wir bereits oben erwähnten - unseren Versuchspersonen insgesamt 94 items an (vgl. Anhang 355 ff.), aus denen dann nach Berechnung der Schwierigkeitsindices diejenigen items ausgewählt wurden, welche bezüglich der Schwierigkeitsindices und der geprüften Lernzielkategorie¹⁾ zu direkt vergleichbaren Testformen führten. Der Berechnung der Schwierigkeitsindices für die Prüfungen "Radizieren" bzw. "Logarithmieren" wurde folgende Formel zugrunde gelegt:

$$P = \frac{N_{R\text{-Gruppe A}} + N_{R\text{-Gruppe B}}}{N_A + N_B}$$

(dies entspricht der allgemeinen Formel $P = \frac{N_R}{N}$)

Dabei bedeuten:

$N_{R\text{-Gruppe A}}$: Summe der Gewichtungspunkte der Gruppe A

$N_{R\text{-Gruppe B}}$: Summe der Gewichtungspunkte der Gruppe B

N_A : Anzahl der maximal erreichbaren Punkte der Gruppe A

N_B : Anzahl der maximal erreichbaren Punkte der Gruppe B

$N_A + N_B$ entspricht der verrechneten Stichprobengröße $N = 119$, da für jedes item eine optimale Gewichtung von 1,0 möglich war.

1) Bei der Konstruktion der ursprünglichen 150 items gingen wir davon aus, daß items, welche leicht zu lösen sind, nicht der Lernzielkategorie "comprehension" zugehören können, da diese mehr Lösungsvoraussetzungen und deren Koordination erfordern. Bei der empirischen Überprüfung der 94 in die den Schülern vorgelegten Testformen aufgenommenen items zeigte sich allerdings vereinzelt, daß items, von denen wir aufgrund unserer bisherigen Informationen annehmen mußten, daß sie schwer zu lösen seien, von den meisten Schülern gelöst worden sind. Items mit solch einem überraschenden empirischen "Effekt" wurden grundsätzlich in keine unserer endgültig verrechneten item-Gruppen aufgenommen.

Mit Hilfe dieser Auslesetechnik gelang es uns, zwei direkt vergleichbare Tests im Sinne unseres Versuchsplans zusammenzustellen. Trotzdem sind die Mittelwerte der Rohpunkte der jeweils miteinander verglichenen Tests (also der Tests "knowledge-Radizieren" und "knowledge-Logarithmieren" bzw. der Tests "comprehension-Radizieren" und "comprehension-Logarithmieren") nicht absolut identisch. Dies liegt daran, daß es nicht gelang, Tests zusammenzustellen, welche bezüglich ihres mittleren Schwierigkeitsindex absolut übereinstimmen. Die vorhandenen, sehr geringen Differenzen der arithmetischen Mittel der Rohwerte sind mittels der abweichenden mittleren Schwierigkeitsindizes erklärbar¹⁾. Wir werden am entsprechenden Ort die mögliche Verfälschung unserer Versuchsergebnisse durch diese mittleren Rohwertsdifferenzen erwähnen und die entsprechenden Korrekturwerte einsetzen.

7.816 Testvorgabe:

Um das Abschreiben während der Testbearbeitung möglichst einzuschränken, wurde die Reihenfolge der items in zwei Variationen vorgegeben²⁾. Darüber hinaus wurden bei einigen items die Buchstaben vertauscht (z.B. wurde anstelle des Ausdrucks $\log_a b$ der Ausdruck $\log_p q$ gewählt). Die Schüler bearbeiteten insgesamt 94 items: ³⁾

1) Die Abweichung der arithmetischen Mittel der Rohwerte beträgt bei den Tests zur Erfassung der Lernzielkategorie "knowledge" $14,515 - 14,375 = 0,140$. Die Differenz der mittleren Schwierigkeitsindizes beträgt hier 0,0005. Auf die 20 items umgerechnet bedeutet dies, daß das arithmetische Mittel des etwas "schwierigen" Tests um ein Hundertstel ($0,0005 \times 20$) tiefer

liegen muß als das arithmetische Mittel des etwas "leichteren" Tests. $14,515 : 100 = 0,145$, dies entspricht tatsächlich der oben festgestellten Differenz der arithmetischen Mittel- und Rohpunktwerte. Die geringfügige Verschiebung bei der dritten Stelle hinter dem Komma muß auf Ungenauigkeiten zurückgeführt werden, welche aufgrund der "Auf-" bzw. "Abrundung" entstanden.

Entsprechend kann die Differenz der arithmetischen Mittel der Rohpunktwerte bei den Tests zur Erfassung der Lernzielkategorie "comprehension" erklärt werden. Hier beträgt die Differenz $3,8007 - 3,6466 = 0,1561$. Die Differenz der mittleren Schwierigkeitsindizes beträgt hier $0,0027$. Auf 15 items umgelegt ergibt sich hier analog zu oben der Wert $0,0405$, d.h. daß das arithmetische Mittel der etwas "schwierigeren" Prüfung um zirka vier Hundertstel tiefer liegen muß als das arithmetische Mittel des etwas "leichteren" Tests. $3,8007 : 100 = 0,0380$ (abgerundet). Multipliziert mit 4 ergibt sich der Wert $0,152$, dieser entspricht der tatsächlich festgestellten Differenz der arithmetischen Mittel der Rohpunktwerte, wobei die Nichtübereinstimmung in der dritten Stelle hinter dem Komma wiederum auf Ungenauigkeiten zurückgeführt werden muß, welche aufgrund von "Auf-" und "Abrundungen" entstanden sind. Die Differenzen der arithmetischen Mittel ($0,145$ bzw. $0,152$) müssen bei der Berechnung der Hypothesen als Korrekturwerte eingesetzt werden.

- 2) Die Reihenfolge wurde grundsätzlich nur bei den ersten items der einzelnen Tests variiert, da nicht ausgeschlossen werden kann, daß die Reihenfolge der Bearbeitung einen Einfluß auf die erreichte Punktzahl hat.
- 3) Die zur Berechnung unserer Hypothesen herangezogenen items finden sich in diesem Kapitel. Die den Schülern vorgelegten Tests finden sich im Anhang.

Nr.	Aufgabe	Lösung	Punktw.
1.	Forme in ein Produkt um: $q^3 p^6 + q^4 p^3 - q^5 p^2$	$q^3 p^2 (p^4 + qp - q^2)$	1
2.	Fasse zusammen: $q^3 p^4 \cdot q^e - 3 p^e - 5$	$q^e \cdot p^{e-1}$	1
3.	Fasse zusammen: $-3q^3 p^2 + 5q^2 p^3 + 8q^2 p^3 + 5q^3 p^2$	$2q^3 p^2 + 3q^2 p^3$	1
4.	Vereinfache den Ausdruck: $\frac{y^q p - 1}{y^q p - 4}$	y^3	1
5.	Vereinfache den Ausdruck: $\frac{y^{q-1} \cdot x^{p-1}}{y^q \cdot x^p}$	$\frac{1}{xy}$	1
6.	Vereinfache soweit als möglich: $\left(\frac{2m}{3n}\right)^a \cdot \left(\frac{9n}{10m}\right)^a$	$\left(\frac{3}{5}\right)^a$	1
7.	Dividiere: $\frac{(q^5 p^2)^y}{(q^6 p)^y}$	$\left(\frac{p}{q}\right)^y$	1
8.	Berechne folgenden Ausdruck soweit als möglich: $(3q^{-2} p^{-3} t^0)^{-3}$	$27q^6 p^9$	1
9.	Rechne folgenden Ausdruck aus: $\sqrt[3]{343}$	7	1
10.	Schreibe als Wurzel: $16^{1/2}$	$\sqrt[2]{16}$	1

VORPRÜFUNG

Nr.	Aufgabe	Lösung	Punktw.
11.	Schreibe als Bruchpotenz: $\sqrt[3]{q}$	$q^{\frac{1}{3}}$	1
12.	Schreibe logarithmisch: $2^x = 16$	$x = \log_2 16$	1
13.	Schreibe in Potenzform: $\log_3 x = 4$	$3^4 = x$	1
14.	Schreibe in Wurzelform: $\log_x 27 = 3$	$x = \sqrt[3]{27}$	1
15.	Beseitige die Brüche: $\frac{a+1}{(1+x)^{a+2}}$	$(a+1)(1+x)^{-a-2}$	1
16.	Vereinfache soweit als möglich: $\frac{(y+5)^{-q-2}}{(y^2+5y)^{-q-2}}$	y^{2+q}	1
17.	Zerlege in Faktoren: $12a^3q^{-2} + 8a^4q^{-3}$	$12a^3q \cdot 12a^{-2} + 8a^4 \cdot 8a^{-3}$	1
18.	Zerlege Zähler und Nenner in Faktoren und kürze soweit als möglich: $\frac{25x^2 - 4x^4}{5x^4 - 2x^5}$	$\frac{2x+5}{x^2}$	1

Prüfungsaufgaben
Wurzelrechnen knowledge

Nr.	P.	Aufgabe	Lösung	Punktw.
1	0,73	Schreibe als Wurzel und berechne gegebenenfalls: $4^{1,5}$	8	1
2	0,74	Berechne die Wurzel u. mache die Probe durch Potenzieren: $\sqrt[3]{125}$	5	1
3	0,86	Schreibe als Wurzel: $y^{-1/4}$	$\frac{1}{\sqrt[4]{y}}$	1
4	0,61	Schreibe als Wurzel: $3^{-0,8}$	$\frac{1}{\sqrt[5]{3^4}}$	1
5	0,63	Schreibe als Wurzel: $\left(\frac{p}{q}\right)^{-1/2}$	$\frac{q}{p}$	1
6	0,83	Schreibe mit gebrochener Hochzahl: $\frac{1}{\sqrt[3]{m}}$	$m^{-1/3}$	1
7	0,69	Schreibe mit gebrochener Hochzahl: $\frac{1}{\sqrt[q]{y}}$	$\frac{y}{q^{\frac{1}{2}}}$	1
8	0,85	Fasse soweit als möglich zusammen: $5^4\sqrt{q} + 3^4\sqrt{q} + 2^3\sqrt{q}$	$8^4\sqrt{q} + 2^3\sqrt{q}$	1
9	0,66	Löse durch Zerlegung: $\sqrt{4000}$	$20\sqrt{10}$	1
10	0,80	Vereinfache soweit als möglich: $\sqrt{\frac{p}{q}} \cdot \sqrt{\frac{q}{z}} \cdot \sqrt{\frac{z}{p}}$	1	1

Nr.	P	Aufgabe	Lösung	Punktw.
11	0,92	Welchen Wert hat die Wurzel: $\frac{14}{\sqrt{0}}$	0	1
12	0,73	Vereinfache soweit als möglich: $\left(\frac{q}{\sqrt{b^p}}\right)^{qm}$	b^{pm}	1
13	0,60	Löse: $\sqrt{16^5}$	1024	1
14	0,84	Vereinfache: $\frac{12}{\sqrt{5^8}}$	$\frac{3}{\sqrt{5^2}}$	1
15	0,57	Ziehe schriftlich die Wurzel: $\sqrt{409600}$	640	1
16	0,84	Dividiere und vereinfache soweit als möglich: $\frac{\sqrt[3]{125}}{\sqrt[3]{216}}$	$\frac{5}{6}$	1
17	0,69	Dividiere und vereinfache soweit als möglich: $\frac{\sqrt[q]{m^2n}}{\sqrt[q]{mn}}$	$\sqrt[q]{m}$	1
18	0,58	Drücke als eine Wurzel aus: $(m+n)\sqrt{\frac{(m-n)}{(m+n)}}$	$\sqrt{m^2-n^2}$	1
19	0,68	Vereinfache soweit als möglich: $\sqrt[3]{\frac{2}{3}}$	$\sqrt[6]{6}$	1
20	0,60	Vereinfache soweit als möglich: $\sqrt{3\sqrt{5}\cdot\sqrt{3}}$	3	1

Prüfungsaufgaben
Wurzelrechnen comprehension

Nr.	p	Aufgabe	Lösung	Punktw.
1	0,12	Vereinfache soweit als möglich: $\frac{(a-b)}{\sqrt{a}-\sqrt{b}}$	$\sqrt{a}+\sqrt{b}$	1
2	0,24	Multipliziere: $(\sqrt[7]{p^2}-\sqrt[7]{pq}+\sqrt[7]{p^2})(\sqrt[7]{p}+\sqrt[7]{q})$	$\sqrt[7]{p^2q}-\sqrt[7]{pq^2}+2\sqrt[7]{p^3}$	1
3	0,24	Fasse soweit als möglich zusammen: $\frac{12}{5}\sqrt[3]{t-z} - \frac{3}{8}\sqrt[3]{t-q}$	$\frac{3}{8}(4\sqrt[3]{t-z} - \sqrt[3]{t-q})$	1
4	0,13	Vereinfache: $2^5\sqrt{4q+4}-4^5\sqrt{9q+9}+3^5\sqrt{q+1}$	$5\sqrt{q+1}(2^5\sqrt{4}-4^5\sqrt{9}+3)$	1
5	0,39	Ziehe soweit als möglich die Wurzel: $\sqrt{18q^3p^6-45q^4p^5}$	$3qp^2\sqrt{2p^2q-5pq^2}$	1
6	0,30	Fasse soweit als möglich zusammen: $(\frac{2}{5}\sqrt{m}-\frac{3}{4}\sqrt{2}) - [\frac{2}{3}\sqrt{m} - (\frac{3}{4}\sqrt{2} + \frac{1}{2}\sqrt{m})]$	$(-\sqrt{m}) \quad \frac{37}{30}\sqrt{m}$	1
7	0,27	Vereinfache soweit als möglich: $\frac{2a}{\sqrt{m}a_y-a} \cdot \frac{2a}{\sqrt{m}3a_y5a}$	m^2y^2	1
8	0,39	Vereinfache soweit als möglich: $4\sqrt{27x^5} \cdot 4\sqrt{3x^3}$	$3x^2$	1
9	0,55	Vereinfache: $\sqrt{8p} \cdot \sqrt{2p^5}$	$4p^3$	1
10	0,14	Vereinfache soweit als möglich: $\frac{x-2\sqrt{xy}+y}{\sqrt{x}-\sqrt{y}}$	$-\sqrt{x}-\sqrt{y}$	1

[illegible]

Prüfungsaufgaben
Logarithmen knowledge

Nr.	p	Aufgabe	Lösung	Punktw.
1	0,83	Schreibe in Potenzform und bestimme den Numerus: $\log_4 x = 3$	$4^3 = x$ $x = 64$	1
2	0,76	Berechne: $\log_5 \frac{1}{625}$	- 4	1
3	0,62	Berechne: $\log_a \frac{1}{4\sqrt{a^4}}$	- 1	1
4	0,68	Berechne: $\log_{1/4} 0,25$	1	1
5	0,70	Schreibe in logarithmischer Form u. bestimme den Exponenten: $\sqrt[3]{1000} = 10$	$x = 3$	1
6	0,66	Schreibe in logarithmischer Form u. bestimme den Exponenten: $4^x = \frac{1}{64}$	$x = \log_4 \frac{1}{64}$	1
7	0,85	Zerlege in Summen und berechne: $\log_{10}(10 \cdot 1\,000 \cdot 10\,000\,000)$	$\log_{10} 10 + \log_{10} 1\,000 + \log_{10} 10\,000\,000 = 11$	1
8	0,86	Zerlege in Differenzen und berechne: $\log_2 \frac{32}{128}$	$\log_2 32 - \log_2 128 = -2$	1
9	0,81	Fasse zusammen und berechne: $\log_3 189 - \log_3 7$	$\log_3 27$ $x = 3$	1
10	0,80	Fasse zusammen und berechne: $\log_2 3 + \log_2 2^{2/3} + \log_2 4$	$\log_2 8$ $x = 3$	1

Nr.	P	Aufgabe	Lösung	Punktw.
11	0,68	Fasse zusammen und berechne: $\log_n(q^2 p^2) - \log_n(q+p)$	$\log_n(q-p)$	1
12	0,82	Bringe in Potenzform und berechne dann: $\log_x 8 = 1/2$	$x = 64$	1
13	0,68	Bringe in Potenzform und berechne: $\log_x 144 = 2$	$x = 12$	1
14	0,73	Stelle als Produkt dar: $\log_a \sqrt[q]{p}$	$\frac{1}{q} \log_a p$	1
15	0,73	Stelle als Produkt dar: $\log_b 10^4$	$4 \cdot \log_b 10$	1
16	0,73	Stelle als Produkt dar: $\log_a p^q$	$q \cdot \log_a p$	1
17	0,69	Stelle als Produkt dar: $\log_a \sqrt[4]{10}$	$1/4 \log_a 10$	1
18	0,53	Zerlege soweit als möglich: $\log_a \frac{q \cdot m \sqrt{q}}{a^2 \cdot z^3}$	$\log_a q + 1/m \cdot \log_a q -$ $2 \log_a a + 3 \log_a z$	1
19	0,70	Rechne soweit als möglich: $\log_{\sqrt{5}} 25$	4	1
20	0,58	Begründe: $\log(u-v) \neq \log u - \log v$	$\log u - \log v = \log \frac{u}{v}$	1

Prüfungsaufgaben
Logarithmen comprehension

Nr.	p	Aufgabe	Lösung	Punktw.
1	0,45	Bringe in Potenzform und berechne: $\log_x 1/2 = -1/3$	$x = 8$	1
2	0,30	Schreibe als Logarithmus einer Zahl: $3 + \log_a 5$	$\log_a 5a^3$	1
3	0,25	Schreibe als Logarithmus einer Zahl: $4 - \log_a 2$	$\log_a \frac{a^4}{2}$	1
4	0,12	Berechne y: $\log y = 1/2 \log(u^2 - v^2) - \log(u+v)$	$y = \sqrt{\frac{u-v}{u+v}}$	1
5	0,21	Berechne y: $\log y = 2 \log(u-v) + 1/2 \log(u+v)$	$y = (u-v)^2 \cdot \sqrt{u+v}$	1
6	0,10	Berechne y: $1/2 \log y \cdot 2 \log(u+v) + 1/2 \log u$	$y = u \cdot (u+v)^4$	1
7	0,25	Zerlege soweit als möglich: $\log \sqrt[3]{\frac{u^2 v^3}{u+v}} = 0,5^{1/2}$	$1/3 [2 \log u + 3 \log v - 1/2 \log 0,5 - \log(u+v)]$	1
8	0,11	Berechne soweit als möglich: $\log_{\sqrt{m}} m \cdot \sqrt{m \sqrt{m}} = x$	$x = 7/2$	1
9	0,08	Berechne soweit als möglich: $\log(x-1) - \log x = 3 \log m - 2 \log mn^{1/2}$	$x = \frac{n}{n-m^2}$	1
10	0,05	Logarithmiere: $\frac{(\frac{1}{u} + \frac{1}{v})(u-v)}{(\frac{1}{u} - \frac{1}{v})(u+v)}$	$\log_{10}(u-v) - \log_{10}(v-u)$	1

7.82 Adressatenbezogene Maßkriterien

Hierunter verstehen wir Kriterien, die nicht unmittelbar auf ein bestimmtes Lehrprogramm bezogen sind, sondern allgemeine Lernvoraussetzungen, Einstellungen und Verhaltensweisen unserer Probanden und der an der Untersuchung beteiligten Lehrer kennzeichnen.

7.821 Intelligenz:

Die Intelligenz der Schüler wurde mit Hilfe des Intelligenzstrukturtests von AMTHAUER gemessen¹⁾. Verrechnet wurde SW-Gesamt, d.h. der Standardwert, welcher aus der Summe sämtlicher Rohwerte gebildet wird.

7.822 "Einstellung zur Schule":

Zur Erfassung der Einstellung unserer Versuchspersonen zur Schule verwendeten wir den Subtest 1 des "Problemfragebogens für Jugendliche" von SÜLLWOLD²⁾. Es handelt sich dabei um ein standardisiertes, bereits auf dem Markt befindliches Verfahren, das speziell für den Altersbereich unserer Stichprobe entwickelt wurde. Der von uns ausgewählte Subtest versucht zu erfassen, inwieweit die Schule und die damit zusammenhängenden Gegebenheiten als problemhaltig erlebt werden. Der Subtest umfaßt 40 items. Die Schüler haben dabei die Auf-

1) R. AMTHAUER: "Intelligenz-Struktur-Test", Göttingen 1955

2) F. SÜLLWOLD: "Problemfragebogen für Jugendliche", Göttingen 1967

gabe, bei jedem der vorgegebenen items zwischen den Gewichtungen 0, 1, 2 und 3 zu wählen. Dabei bedeutet die Gewichtung 0, daß der Schüler den in dem betreffenden item angesprochenen Sachverhalt als nicht problemhaltig erlebt, die Gewichtung 3 bedeutet, daß der angesprochene Sachverhalt als ausgesprochen problemhaltig erlebt wird. Den Korrelationsberechnungen wurde die Summe aller vorgenommenen Gewichtungen zugrunde gelegt.

7.823 "Einstellung zum Lehrer":

Die Einstellung unserer Versuchspersonen zu den im Versuch unterrichtenden Lehrern wurde mittels numerischer Schätzskalen festgestellt¹⁾. Diese waren fünfstufig angelegt und umfaßten 9 polar angeordnete Qualitäten, welche in Form von Eigenschaftsworten formuliert waren. Die fünf Abstufungen waren verbal gekennzeichnet²⁾. Die Schüler wurden aufgefordert, unabhängig voneinander die entsprechenden Wertungen vorzunehmen, wobei sie ausdrücklich die Versicherung erhielten, daß der Lehrer keine Einsicht in die Testformulare bekäme. Die Lehrer waren zum Zeitpunkt der Bewertung nicht anwesend. Die Messung wurde unmittelbar vor Beginn der experimentellen Phase vorgenommen. Den Korrelationsberechnungen wurde die Summe aller neun Gewichtungen zugrunde gelegt.

7.824 "Konzentrativer Arbeitshaltung":

Die an der Untersuchung beteiligten Lehrer bekamen die Aufgabe, jeden einzelnen Schüler im Hinblick auf seine

- 1) Entsprechende Verfahren sind in der Unterrichtsforschung außerordentlich verbreitet (vgl. L. TENT: "Schätzverfahren in der Unterrichtsforschung", in: K.-H. INGENKAMP: "Handbuch der Unterrichtsforschung Teil I", Weinheim-Berlin-Basel 1970, S. 864). Dies liegt wohl daran, daß es verhältnismäßig einfach ist, "numerische Schätzskalen zu konstruieren und anzuwenden. Auch die Weiterverarbeitung der damit gewonnenen Daten ist verhältnismäßig einfach" (L. TENT, a.a.O., S. 869). Die Frage, ob es sich bei numerischen Schätzskalen um Rang- oder um Intervallinformationen handelt, wird unterschiedlich beurteilt. Obwohl streng genommen gleiche Intervallabstände mittels psychophysischer Untersuchungen erst nachgewiesen werden müßten, wird häufig auf eine entsprechende Überprüfung verzichtet und auf Intervallskalenniveau verrechnet. Vereinzelt wird sogar die Ansicht vertreten, daß "Schätzurteile selbst dann die Eigenschaften von Intervall- und Verhältnisskalen besitzen, wenn den Kategorien keine Zahlen zugeordnet werden" (L. TENT, a.a.O., S. 870).
- 2) Die Eigenschaftspaare wurden einer Veröffentlichung von THOMAE entnommen (vgl. H. THOMAE: "Das Individuum und seine Welt. Eine Persönlichkeitstheorie", Göttingen 1968, S. 254) und umfassen folgende Qualitäten: "Freundlich - feindlich, verständnisvoll - verständnislos, zugewandt - abgewandt, versprechend - drohend, gewährend - versagend, offen - verschlossen, vertraut - fremd, höflich - unhöflich, sicher - unsicher". Eine an einer unseren Versuchspersonen vergleichbaren Stichprobe von $N = 20$ vorgenommene Überprüfung der Stabilität des Meßinstruments ergab einen Korrelationskoeffizienten von $r = 0,93$. Zu diesem Zweck wurden 20 Schüler aufgefordert, ihren Klassenlehrer zu beurteilen. Die Messung wurde nach Ablauf von drei Tagen wiederholt, wobei die Schüler ausdrücklich darauf hingewiesen wurden, daß es nicht erforderlich sei, daß sie sich an die bei der ersten Messung angestrichenen Wertungen erinnern. Die Schüler hatten in dem zwischen den beiden Wertungen liegenden Zeitraum keinen Kontakt zu dem beurteilten Lehrer, so daß eine Beeinflussung der beiden Werte durch zwischenzeitlich eingetretene Erlebnisse mit der beurteilten Person ausgeschlossen war.

"konzentrativer Arbeitshaltung" einzustufen. Darunter verstehen wir die Fähigkeit des Schülers, ohne Hilfe und Zuspruch des Lehrers selbständig, konzentriert und kontinuierlich zu arbeiten. Zur Erfassung dieses Aspektes des Schülerverhaltens verwendeten wir wiederum fünfstufige numerische Schätzskalen. Sie umfaßten drei polar angeordnete Schülerverhaltensweisen¹⁾. Die Wertungen wurden unmittelbar nach Abschluß der experimentellen Phase vorgenommen. Den Korrelationsrechnungen wurde die Summe aller drei Gewichtungungen zugrunde gelegt.

-
- 1) Die 3 polar angeordneten Schülerverhaltensweisen wurden von 9 Lehrerstudenten aus einer vorgegebenen Liste, welche 8 Wahlmöglichkeiten enthielt, ausgewählt. Sie erhielten dabei die Instruktion, diejenigen drei Paare anzustreichen, welche nach ihrer Auffassung am besten geeignet sind, um selbständiges, konzentriertes und kontinuierliches Arbeitsverhalten zu kennzeichnen. Dabei wurden folgende polar angeordnete Schülerverhaltensweisen am häufigsten gewählt: "Der Schüler kann sich nicht längere Zeit allein mit einem Stoff beschäftigen - der Schüler kann sich längere Zeit allein mit einem Stoff beschäftigen; der Schüler schweift leicht ab - der Schüler schweift nicht ab; der Schüler gibt bei Schwierigkeiten leicht auf - der Schüler gibt bei Schwierigkeiten nicht auf." Zur Überprüfung der Stabilität unseres Meßinstruments wurden drei Lehrer, welche an derselben (unserer Stichprobe vergleichbaren) Klasse unterrichteten, aufgefordert, die entsprechenden Wertungen vorzunehmen. Dabei ergab sich ein mittlerer Korrelationskoeffizient von $r = 0,83$ ($N = 27$). Dieser mittlere Korrelationskoeffizient kann als Hinweis dafür dienen, daß die polar angeordneten Schülerverhaltensweisen so eindeutig definiert sind, daß keine massiven Beobachterdiskrepanzen entstehen, welche auf eine unzulängliche Definition der Beurteilungskriterien zurückzuführen wären.

7.825 Positive Verstärkungen und "stoffbezogene Schüleraktivität"

Die hier angesprochenen adressatenbezogenen Maßkriterien wurden unter Punkt 7.72 und 7.73 ausführlich behandelt. An dieser Stelle sei lediglich noch einmal ausdrücklich darauf hingewiesen, daß unter dem Begriff "stoffbezogene Schüleraktivität" keine "Eigenschaft" sondern lediglich die Häufigkeit des Auftretens definierten Schülerverhaltens verstanden wird.

7.90 Abschließende Diskussion methodischer Probleme

Wir haben unter Punkt 7.23 die für Vergleichsuntersuchungen wesentlichen Fehlerquellen zusammenfassend aufgeführt. Im folgenden soll unter Bezugnahme auf die einzelnen Punkte von 7.23 unsere methodische Konzeption abschließend diskutiert werden.

- a) Die experimentelle Phase betrug einschließlich der Prüfungen pro Klasse zirka 15 Unterrichtsstunden¹⁾, die Untersuchung lief in den einzelnen Klassen über zirka 3 Wochen. Dies ist, gemessen an "Laboruntersuchungen", zweifellos ein langer Zeitraum, so daß wir mit Störeffekten, welche durch zwischenzeitliches Geschehen verursacht werden, rechnen müssen. Es ist jedoch sehr unwahrscheinlich, daß sich solchermaßen entstandene Störvarianz im Rahmen unseres Kreuzvalidierungskonzepts bei allen Vergleichs-

1) Von einer Schulstunde wurden nur 20 Minuten beansprucht (Prüfung "Radizieren", 3. Teil).

richtungen in gleicher Weise auswirkt und somit unsere Untersuchungsergebnisse systematisch verzerrt. Ungeachtet dieser Tatsache meinen wir jedoch darauf hinweisen zu müssen, daß die Wahl einer deutlich geringeren Versuchsdauer die Gefahr von zwischenzeitlich verursachter systematischer Fehlervarianz zwar vermindert hätte, daß gleichzeitig jedoch auch der Nachweis systematischer Beeinflussung der Schülerleistung durch die verschiedenen Unterrichtsmethoden beeinträchtigt worden wäre. Der Rückgriff auf Doppelstunden schien uns nicht sinnvoll zu sein, da hier zweifellos die Gefahr der Überbeanspruchung der Versuchspersonen bestanden hätte. Dies gilt insbesondere für die Programmgruppen, da aufgrund der einschlägigen Erfahrungen im Umgang mit programmiertem Unterrichtsmaterial von einer sehr hohen konzentrativen Beanspruchung ausgegangen werden muß¹⁾.

- b) Dadurch, daß der Test zur Ermittlung des Vorwissens mit den Tests, mit deren Hilfe der Effekt der Versuchsunterrichte gemessen wurde, nicht identisch ist, konnte eine Störvarianz, welche durch die Reaktivität der Meßinstrumente verursacht wird, nicht entstehen.
- c) Eine absolut objektive Auswertung der Tests, mittels derer unterrichtsbezogene Maßkriterien gemessen worden wären, wäre zwar formal möglich gewesen, dies hätte aber einen Verzicht auf die eigentliche Intention von kriteriumsbezogenen Tests bedeutet, da der Aspekt der "teilweisen Erreichung" eines Lern-

1) Vgl. Punkt 2.30

ziels außer acht gelassen worden wäre. Im übrigen können wir trotz der von uns gewählten abgestuften, nicht ganz objektiven Gewichtung von einer hohen Objektivität der Auswertung ausgehen, da die Tests von zwei unabhängigen Auswertern korrigiert worden sind und im (seltenen) Fall abweichender Gewichtung einzelner items eine gemeinsame Entscheidung nach sorgfältiger Analyse getroffen worden ist¹⁾. Die von uns erhobenen Beobachtungsdaten können, jedenfalls soweit sie im Rahmen der Prüfung unserer Hypothesen herangezogen wurden, aufgrund des von uns gewählten Kriteriums "vorhanden" bzw. "nicht vorhanden" und dem Verzicht auf qualitative Abstufungen als zuverlässig betrachtet werden. Dies wurde durch die Überprüfung der Beobachterübereinstimmung bestätigt (vgl. 7.73)

- d) Das Problem der statistischen Regression stellt sich im Rahmen des Berichts unserer Versuchsergebnisse dort, wo wir auf Extremgruppen zurückgreifen.

1) Die Abweichungen der Gewichtung durch die beiden Korrektoren betrug nie mehr als 0,25 Punkte. In der überwiegenden Anzahl der Fälle war es möglich, nach erfolgter gemeinsamer Analyse sich begründet gemeinsam für eine Gewichtungsstufe zu entscheiden. Lediglich in 17 Fällen war es uns nicht möglich, den Gedankengang der betreffenden Schüler zu rekonstruieren, so daß wir uns hier "willkürlich" für die geringere der beiden Gewichtungsmöglichkeiten entschieden. Angesichts der Tatsache, daß 119 Schüler je 70 items bearbeitet haben, fallen diese 17 Fälle so gut wie nicht ins Gewicht (es handelt sich lediglich um zirka 2 Tausendstel der insgesamt zu korrigierenden items!).

Wir werden an entsprechender Stelle darauf zurückkommen.

- e) Ohne Randomisierung muß immer damit gerechnet werden, daß aufgrund abweichender Stichprobenzusammensetzung systematische Verzerrungen der abhängigen Variablen möglich sind. Dasselbe gilt für die Problematik der Verzerrung von Ergebnissen infolge experimenteller Einbußen, da diese im Rahmen von längerfristig angesetzten Vergleichsstudien nicht vermieden werden können. Wir gehen allerdings davon aus, daß wir durch die Wahl unseres Versuchsplans ein für natürliche Klasseneinheiten optimales Konzept gewählt haben. Im Falle des Auftretens systematischer Störvarianz aufgrund abweichender Stichproben bei dem Vergleich von Feld 1 versus 2 bzw. Feld 3 versus 4 müßte diese in beiden Fällen in der gleichen Richtung wirksam werden, um durchgehend einen Methodeneffekt der Programmierten Unterweisung vorzutauschen. Dies ist immerhin deutlich unwahrscheinlicher als im Falle der Verwendung einer einfachen quasiexperimentellen Anordnung mit nicht gleichartiger Kontrollgruppe. Wechselwirkungseffekte von Stichprobe, Situation und Meßinstrument sind zwar denkbar, aber angesichts der Tatsache, daß wir von nachgewiesenen Methodeneffekten erst dann ausgehen, wenn wenigstens beim Vergleich der Felder 1 versus 4 und 3 versus 2 signifikante Unterschiede in derselben Richtung auftreten und diese möglichst durch den Vergleich der Felder 1 versus 2 und 3 versus 4 bestätigt werden sollten, ist es so gut wie ausgeschlossen, daß die Wechselwirkungen immer derart gerichtet sind, daß sie eine höhere Lehreffektivität der Programmierten Unterweisung durchgehend vorzutauschen vermögen.

- f) Die Mitarbeit der Schulen erfolgte nicht auf freiwilliger Basis. Durch die seitens der Schulbehörde gelenkte Zuteilung wurde die Fehlerquelle, die durch die Wahl von Schulen mit besonderer Aufgeschlossenheit gegenüber neuen Methoden verursacht wird, eingeschränkt.
- g) Der Verfälschung unserer Versuchsergebnisse zugunsten der Programmierten Unterweisung aufgrund einschlägiger Schülerkenntnisse wurde dadurch entgegengewirkt, daß den Schülern die Erwartungen des Versuchsleiters nicht bekannt waren. Im übrigen muß in diesem Zusammenhang auf die Messungen der Einstellungen der Schüler zur Programmierten Unterweisung und zum konventionellen Lehrerunterricht hingewiesen werden, welche bei beiden Gruppen eine hochsignifikante Bevorzugung des Lehrerunterrichts ergaben¹⁾. Allein angesichts dieser Ergebnisse kann nicht angenommen werden, daß die Schüler im Lehrerunterricht absichtlich schlechter mitgearbeitet haben und dadurch Verfälschungen zugunsten der Programmierten Unterweisung entstehen konnten.
- h) Der insbesondere aufgrund der Beobachtung erforderliche organisatorische und technische Aufwand war für die Versuchspersonen deutlich erkennbar. Dem wurde durch die Gewöhnungsphase und der Platzierung der Beobachter im Rücken der Schüler entgegengewirkt. Weiter wurden die reaktiven Effekte der experimentellen Situation durch den Rückgriff auf die stundenplanmäßig vorgesehenen Mathematikstunden, die Beibehaltung der gewohnten räumlichen Umgebung, den Erhalt der natürlichen Klasseneinheiten und den Einsatz der regulären Mathematiklehrer eingeschränkt.

- 1) Zur Feststellung der Einstellung gegenüber dem Lehrerunterricht bzw. der Programmierten Unterweisung wurde unmittelbar nach Abschluß der Untersuchung eine Einstellungsmessung mittels Ratingskalen vorgenommen. Diese waren 5-stufig angeordnet und umfaßten 6 Alternativmerkmale. 1. "macht Spaß" - "macht keinen Spaß"; 2. "interessant" - "langweilig"; 3. "nicht anstrengend" - "anstrengend"; 4. "verständlich" - "unverständlich"; 5. "ich möchte häufig so unterrichtet werden" - "ich möchte nie mehr so unterrichtet werden"; 6. "bereitete gut auf Prüfungen vor" - "bereitete schlecht auf Prüfungen vor". Die Reihenfolge der Beurteilung wurde systematisch variiert, für beide Unterrichte wurden dieselben Ratingskalen verwendet. Die fünf Abstufungen waren verbal gekennzeichnet ("sehr" (1), "ziemlich" (2), "unentschieden" (3), "ziemlich" (4), "sehr" (5)). Überraschenderweise ergab sich hier bei beiden Gruppen eine hochsignifikante Bevorzugung des Lehrerunterrichts. Bei Gruppe A fanden wir einen t-Wert von 7,06, bei Gruppe B einen t-Wert von 5,02. Das entspricht einer Auftrittswahrscheinlichkeit von $p < 1\%$. Es spricht einiges dafür, daß die massive Ablehnung der Programmierten Unterweisung auf die gehäuften und z.T. als ausgesprochen schwierig erlebten Prüfungen zurückgeführt werden muß. Offenbar wurden die durch die experimentelle Phase an die Schüler gestellten höheren Anforderungen mit der neuen Lehrmethode assoziiert. Obwohl auch der herkömmliche Lehrerunterricht straffer und mit höherem Anspruch an die Schüleraktivität, als dies üblicherweise der Fall ist, geführt worden ist (dies wurde uns sowohl seitens der Lehrer als auch der Schüler mehrfach bestätigt) konzentrierte sich die negative Einstellung fast ausschließlich auf die Programmierte Unterweisung. Wir weisen ausdrücklich darauf hin, daß von dieser Messung keine Rückschlüsse auf die Einstellung der Schüler gegenüber der Programmierten Unterweisung zulässig sind, sofern diese im Schulalltag verwendet wird, vielmehr muß das Ergebnis unserer Messung primär auf unsere experimentellen Bedingungen zurückgeführt werden. Ungeachtet dieser Einschränkung scheint es uns jedoch notwendig zu sein, darauf hinzuweisen, daß die häufig berichtete durchgehend "begeisterte Aufnahme" der Programmierten Unterweisung durch die Schüler kritisch überprüft werden sollte, da in den einschlägigen Untersuchungen, abgesehen vom "Novitätseffekt" vermutlich noch eine Vielzahl anderer Störvariablen nicht berücksichtigt worden sind.

- i) Die in den Prüfungen verwendeten items entsprechen den Aufgaben der üblichen Schularbeiten. Aus diesem Grund beschränkten wir uns auch auf "free-response-items" und verzichteten auf "Alternativ-response" und "multiple-choice-response-items", obwohl die Experten von Schulleistungstests insbesondere letzteren den Vorzug geben¹⁾. Durch den Hinweis, daß die Prüfungen für die Benotung herangezogen werden, wurde der "Charakter" regulärer Prüfungsbedingungen weiter verfestigt.

Zusammenfassend glauben wir feststellen zu können, daß, ungeachtet der vorhandenen Mängel unseres Versuchsplans, dieser gegenüber der bisherigen Praxis der Vergleichsuntersuchungen eine echte Verbesserung darstellt.

1) Vgl. P. STRITTMATTER, a.a.O., S. 161

8.00 Versuchsergebnisse¹⁾

8.10 Hypothesen, welche die Ausprägung der Lernleistung betreffen

8.11 Normalität der Verteilungen:

Voraussetzung für die unten verwendeten parametrischen Prüfverfahren ist der Nachweis, daß die Annahme der Normalverteilung unserer Maßzahlen beibehalten werden kann. Wir müssen daher unsere empirisch gefundene Maßzahlverteilung mit einer theoretisch unterstellten Verteilung, in unserem Fall der Normalverteilung, vergleichen. Zu diesem Zweck haben wir unsere Stichprobe in je fünf Klassen unterteilt und die theoretisch erwartete Häufigkeit mit der tatsächlich beobachteten Häufigkeit mit Hilfe des χ^2 -Tests auf signifikante Abweichungen hin untersucht²⁾. Der im Rahmen der Unterrichtsforchung häufig beobachtbare Verzicht auf eine solche Überprüfung ist umso bedenklicher, als es sich bei den einschlägigen Untersuchungen (wie auch in unserem Fall) nicht um Zufallsstichproben handelt, d.h. aber, daß die Wahrscheinlichkeit einer Abweichung von der Normalverteilung deutlich höher angesetzt werden muß. Zum anderen handelt es sich bei den im Rahmen solcher Untersuchungen zur Ermittlung der Leistung herangezogenen Verfahren fast ausschließlich um eigens für

1) Die der jeweiligen Verrechnung zugrunde liegenden Maßzahlen finden sich im Anhang dieser Arbeit.

2) Vgl. G.A. LIENERT: "Verteilungsfreie Methoden in der Biostatistik", Meisenheim 1962, S. 74 ff.

die Untersuchung entworfene Meßinstrumente, bei welchen man eine Normalverteilung der von ihnen erfaßten Schülerleistungen nicht mit der Sicherheit erwarten kann, wie das bei erprobten und standardisierten Schulleistungstests der Fall ist. Selbst das Vorliegen einer der beiden oben genannten Gründe genügt, um bei entsprechenden Untersuchungen eine Überprüfung der Normalverteilung für notwendig zu halten¹⁾.

Die für die Gruppen A und B getrennt vorgenommene Überprüfung der Güte der Anpassung an die Normalverteilung ergab folgende χ^2 -Werte:²⁾

$A_{\text{knowl.}}/KU$	$\chi^2 = 0,42$
$A_{\text{knowl.}}/PU$	$\chi^2 = 5,26$
$A_{\text{compr.}}/KU$	$\chi^2 = 7,36$
$A_{\text{compr.}}/PU$	$\chi^2 = 7,51$
$B_{\text{knowl.}}/KU$	$\chi^2 = 1,50$
$B_{\text{knowl.}}/PU$	$\chi^2 = 2,38$
$B_{\text{compr.}}/KU$	$\chi^2 = 2,73$
$B_{\text{compr.}}/PU$	$\chi^2 = 1,67$

$\chi^2_{\text{krit. 5\%}}$ ist für alle 8 Fälle 9,49 (df = 4), d.h. wir müssen die Nullhypothese beibehalten und können davon ausgehen, daß unsere Stichproben jeweils einer normalverteilten Grundgesamtheit entstammen. Die χ^2 -Werte bei den Prüfungen $A_{\text{compr.}}$ liegen allerdings beide ziemlich nahe an $\chi^2_{\text{krit. 5\%}}$, hier handelt es sich um eine schlechte Güte der Anpassung an die Normalverteilung. Etwas eingeschränkt gilt dies auch für den Wert $A_{\text{knowl.}}/PU$, in allen übrigen Fällen kann die Güte der Anpassung als gut bezeichnet werden³⁾.

-
- 1) Eine Reihe von Autoren vertritt allerdings die Ansicht, daß parametrische Verfahren relativ unempfindlich auch gegen grobe Abweichungen von der Normalverteilung seien (vgl. P. HOFSTÄTTER u. D. WENDT: "Quantitative Methoden der Psychologie", München 1967, S. 129). In diesem Falle sollte man sich allerdings nach unserer Ansicht nichtparametrischer Methoden bedienen, zumal der dadurch verursachte Informationsverlust nicht so groß ist, wie mancherorts angenommen wird. Im Zusammenhang mit der Verwendung parametrischer Verfahren muß noch darauf hingewiesen werden, daß einige Autoren davon ausgehen, daß Leistungstests, und um solche handelt es sich in unserem Fall, nur den Informationswert von Rangplätzen haben. In diesem Zusammenhang schreibt LIENERT: "Auch der Psychologe hat keine Gewähr, daß ein Proband, der 7 Aufgaben eines Tests (mit 10 Aufgaben) löst, leistungsmäßig eben soweit über einem Probanden steht, der 6 Aufgaben gelöst hat, wie ein Proband, der 11 Aufgaben gelöst hat, über einem solchen mit 10 Aufgaben steht. Die Maßeinheit "Testaufgabe" ist keine Einheit im Sinne der Intervallskala, da eine Testaufgabe der anderen nie vollkommen gleicht" (G.A. LIENERT, a.a.O., S. 48). Dieser strenge Maßstab wird keinesfalls von allen Autoren angelegt. Abgesehen davon, daß fast alle einschlägigen Untersuchungen von der Annahme eines Intervallskalenniveaus ausgehen (vgl. P. STRITTMATTER: "Präsentationsmodi von Lehrprogrammen", Weinheim 1970, und M. HIRZEL: "Partnerarbeit und Programmierter Unterricht", Stuttgart 1970) und entsprechend mit parametrischen Prüfverfahren arbeiten, zeigt sich auch, daß in statistischen Lehrbüchern Daten aus Leistungstests auf der Basis des Intervallskalenniveaus verrechnet werden (vgl. E. MITTENECKER: "Planung und statistische Auswertung von Experimenten", Wien 1963, S. 147 ff.). Angesichts der üblichen Handhabung der Daten von Leistungstests und im Hinblick darauf, daß sich bezüglich der Normalverteilung der Schwierigkeitsindizes unserer vier Tests eine befriedigende bis gute Güte der Anpassung zeigte (vgl. 7.815), haben wir uns entschlossen, ebenfalls auf parametrische Prüfverfahren zurückzugreifen und damit Intervallskalenniveau vorauszusetzen.
- 2) Nähere Angaben finden sich im Anhang.
- 3) Vgl. G.A. LIENERT, a.a.O., S. 74.

8.12 Varianzen der Lernleistung

a) Die Überprüfung der Varianzhomogenität zwischen den unabhängigen Gruppen A und B erfolgt mittels des F-Tests¹⁾. Dabei ergaben sich folgende Werte:

$$s^2_{A/\text{knowl.}/KU} : s^2_{B/\text{knowl.}/PU} : F = \frac{8,262}{7,411} = \underline{1,11} \text{ (nicht sign.)}$$

$$s^2_{B/\text{knowl.}/KU} : s^2_{A/\text{knowl.}/PU} : F = \frac{13,893}{9,675} = \underline{1,44} \text{ (nicht sign.)}$$

$$s^2_{A/\text{compr.}/PU} : s^2_{B/\text{compr.}/KU} : F = \frac{5,656}{5,250} = \underline{1,08} \text{ (nicht sign.)}$$

$$s^2_{B/\text{compr.}/PU} : s^2_{A/\text{compr.}/KU} : F = \frac{10,786}{4,557} = \underline{2,37} \text{ (nicht sign.)}$$

Für die ersten drei Werte kann die Nullhypothese beibehalten werden, d.h. die Varianzhomogenität muß als gegeben betrachtet werden. Dagegen liegt beim letzten Wert ($F = 2,37$; $df_1 = 56$, $df_2 = 61$) eine auf dem 1%-Niveau gesicherte Abweichung der Varianzen vor, hier muß die Annahme der Varianzhomogenität verworfen werden.

b) Die Überprüfung der Varianzhomogenität korrelierender Stichproben erfolgte mittels der Formel²⁾

$$t = \frac{(s^2_1 - s^2_2) \sqrt{N-2}}{2 \cdot \sqrt{s^2_1 \cdot s^2_2 (1 - r_{12})^2}} \quad (df = N - 2)$$

1) Vgl. E. MITTENECKER, a.a.O., S. 69 ff.

2) Vgl. E. MITTENECKER, a.a.O., S. 127

Dabei bedeuten:

- s^2_1 : Größere Varianz der beiden verglichenen Meßwertreihen
 s^2_2 : Kleinere Varianz der beiden verglichenen Meßwertreihen
 N : Größe der Stichprobe
 r_{12} : Korrelation zwischen den beiden Meßwertreihen

Hier ergaben sich folgende Ergebnisse:¹⁾

$$s^2_{A/\text{knowl.}/KU} : s^2_{A/\text{knowl.}/PU} : \underline{t = 0,678} \quad (t_{\text{krit } 5\%} = 2,000)$$

$$s^2_{A/\text{compr.}/KU} : s^2_{A/\text{compr.}/PU} : \underline{t = 0,897} \quad (t_{\text{krit } 5\%} = 2,000)$$

$$s^2_{B/\text{knowl.}/KU} : s^2_{B/\text{knowl.}/PU} : \underline{t = 2,630} \quad (t_{\text{krit } 5\%} = 2,005)$$

$$s^2_{B/\text{compr.}/PU} : s^2_{B/\text{compr.}/KU} : \underline{t = 3,461} \quad (t_{\text{krit } 1\%} = 2,668)$$

Für das Verhältnis der Varianzen der Gruppe A kann die Nullhypothese nicht verworfen werden, bei der Gruppe B kann dagegen in beiden Fällen nicht mehr von einer Varianzhomogenität ausgegangen werden. Hier weist bei der Lernzielkategorie "knowledge" der herkömmliche Unterricht eine signifikant größere Streuung der Lernleistungen auf, bei der Lernzielkategorie "comprehension" zeigt umgekehrt die Leistung bei Programmierter Unterweisung eine deutlich größere Varianz (1%-Niveau).

-
- 1) Die zur Berechnung des t-Wertes notwendigen Korrelationskoeffizienten wurden mittels der Produkt-Moment-Korrelationstechnik errechnet. Dabei ergaben sich folgende Werte:

$$r_{A/\text{knowl.}/KU - A/\text{knowl.}/PU} : 0,43$$

$$r_{A/\text{compr.}/KU - A/\text{compr.}/PU} : 0,35$$

$$r_{B/\text{knowl.}/KU - B/\text{knowl.}/PU} : 0,43$$

$$r_{B/\text{compr.}/PU - B/\text{compr.}/KU} : 0,61$$

Die den Berechnungen zugrunde gelegten Varianzen finden sich unter 8.12 a).

8.13 Lernerfolg in der Lernzielkategorie "knowledge"

a) Für den Vergleich der Lernleistung im Falle von abhängigen Stichproben verwendeten wir zur Ermittlung des Stichprobenfehlers der Differenzen folgende Formel:¹⁾

$$\sigma_{\text{Diff}} = \sqrt{\frac{\xi d^2}{N(N-1)}} = \sqrt{\frac{\xi D^2 - \frac{(\sum D)^2}{N}}{N(N-1)}}$$

(dabei bedeutet:

|D| : Betrag der Differenz zwischen der Lernleistung im Programmierten und im herkömmlichen Unterricht
(PU - KU)

N : Größe der Stichprobe)

H 1: Vergleich B/knowl./PU - B/knowl./KU:

$$\sigma_{\text{Diff}} = 0,460$$

$$(M_{\text{B/knowl./PU}} = 15,25; M_{\text{B/knowl./KU}} = 14,08) \quad 2)$$

$$t = 2,543 \quad (df = 56)$$

Da der Mittelwert des Tests "Rechnen mit Logarithmen" in der Lernzielkategorie "knowledge" um 0,140 tiefer liegt als im Test "Wurzelrechnen" (vgl. 7.815) und dadurch der t-Wert in Richtung unserer Hypothese unge-rechtfertigt vergrößert würde, wurde die in den t-Wert

1) Vgl. E. MITTENECKER, a.a.O., S. 114 ff.

2) Der Mittelwert der Differenzen D ist identisch mit der Differenz der Mittelwerte der X_1 - und X_2 -Verteilung. Wir haben uns daher an dieser Stelle und im folgenden darauf beschränkt, unsere tatsächlichen Gruppenmittelwerte anzugeben.

eingehende Differenz der Mittelwerte entsprechend korrigiert (d.h. von der Differenz der oben angegebenen Mittelwerte wurde der Wert 0,140 subtrahiert). Entsprechend verringerte sich der t-Wert auf

$$\boxed{t_{\text{kor}} = 2,239} \quad (df = 56) \quad (t_{\text{krit}} 5\% = 2,005) \\ (t_{\text{krit}} 1\% = 2,668)$$

H 2: Vergleich A/knowl./PU - A/knowl./KU:

$$d_{\text{Diff}} = 0,402$$

$$(M_{\text{A/knowl./PU}} = 14,67; M_{\text{A/knowl./KU}} = 13,78)$$

$$\underline{t = 2,214} \quad (df = 61)$$

Hier wurde, da die Gruppe A das Rechnen mit Logarithmen programmiert gelernt hat und der Mittelwert dieser Prüfung in der Lernzielkategorie "knowledge" um 0,140 tiefer liegt als der Testmittelwert der Prüfung "Wurzelrechnen" (vgl. 7.851) und dadurch der t-Wert ungerechtfertigt gegen die Richtung unserer Hypothese beeinflusst würde, die in den t-Wert eingehende Differenz der Mittelwerte um 0,140 vergrößert. Entsprechend vergrößert sich der t-Wert auf

$$\boxed{t_{\text{kor}} = 2,562} \quad (df = 61) \quad (t_{\text{krit}} 5\% = 2,000) \\ (t_{\text{krit}} 1\% = 2,660)$$

b) Für den Vergleich der Lernleistung im Falle von unabhängigen Stichproben verwendeten wir zur Ermittlung des Stichprobenfehlers der Differenzen die Formel¹⁾

1) Vgl. E. MITTENECKER, a.a.O., S. 60 ff.

$$s_{\text{Diff}} = \sqrt{\frac{s_{x_1}^2 + s_{x_2}^2}{N_1 + N_2 - 2} \cdot \left(\frac{1}{N_1} + \frac{1}{N_2} \right)}$$

H 3: Vergleich B/knowl./PU - A/knowl./KU: ¹⁾

$$s_{\text{Diff}} = 0,514$$

$$(M_{\text{B/knowl./PU}} = 15,25; M_{\text{A/knowl./KU}} = 13,78)$$

$$t = 2,860 \quad (df = 117) \quad (t_{\text{krit } 1\%} = 2,626)$$

H 4: Vergleich A/knowl./PU - B/knowl./KU:

$$s_{\text{Diff}} = 0,628$$

$$(M_{\text{A/knowl./PU}} = 14,67; M_{\text{B/knowl./KU}} = 14,08)$$

$$t = 0,939 \quad (df = 117) \quad (t_{\text{krit } 5\%} = 1,984)$$

Zusammenfassung:

H 1: Die Nullhypothese muß verworfen werden

H 2: Die Nullhypothese muß verworfen werden

H 3: Die Nullhypothese muß verworfen werden

H 4: Die Nullhypothese muß beibehalten werden.

1) Bei den Hypothesen 3 und 4 (bzw. 7 und 8) ist eine Korrektur des t-Wertes nicht erforderlich, da hier jeweils Maßzahlen aus demselben Test miteinander verglichen werden.

8.14 Lernerfolg in der Lernzielkategorie "comprehension"

a) Vergleich der Lernleistung abhängiger Stichproben

Zur Verrechnung wurde dieselbe Formel wie unter Punkt 8.13 a) herangezogen.

H 5: Vergleich B/compr./PU - B/compr./KU:

$$\delta_{\text{Diff}} = 0,325$$

$$(M_{\text{B/compr./PU}} = 4,06; M_{\text{B/compr./KU}} = 3,12)$$

$$\underline{t = 2,892} \quad (df = 56) \quad (t_{\text{krit } 1\%} = 2,668)$$

Analog zu Punkt 8.13 a) wurde von der tatsächlichen Differenz der Mittelwerte die Differenz der beiden Testmittelwerte zur Erfassung der Lernzielkategorie "comprehension" subtrahiert, da sonst der t-Wert in Richtung unserer Hypothese ungerechtfertigt vergrößert würde. Es handelt sich dabei um den Wert 0,156 (vgl. 7.851). Dabei ergab sich ein korrigierter t-Wert von

$$\boxed{t_{\text{korr}} = 2,412} \quad (df = 56) \quad (t_{\text{krit } 5\%} = 2,005)$$

H 6: Vergleich A/compr./PU - A/compr./KU:

$$\delta_{\text{Diff}} = 0,370$$

$$(M_{\text{A/compr./PU}} = 4,17; M_{\text{A/compr./KU}} = 3,54)$$

$$\underline{t = 1,703} \quad (df = 61) \quad (t_{\text{krit } 5\%} = 2,000)$$

Analog zu Punkt 8.13 a) wurde zu der tatsächlichen Differenz der Mittelwerte die Differenz zwischen den beiden Testmittelwerten zur Erfassung der Lernzielkategorie "comprehension" addiert, da sonst der t-Wert gegen die Richtung unserer Hypothese ungerechtfertigt verkleinert

würde. Es handelt sich dabei um den Wert 0,156 (vgl. 7.851). Dabei ergab sich ein korrigierter t-Wert von

$$\boxed{t_{\text{korr}} = 2,124} \quad (df = 61) \quad (t_{\text{krit } 5\%} = 2,000)$$

- b) Vergleich der Lernleistung unabhängiger Stichproben

H 7: Vergleich B/compr./PU - A/compr./KU:

Da in diesem Fall bezüglich der Varianzhomogenität die Nullhypothese nicht beibehalten werden konnte¹⁾, können wir nicht von einer gemeinsamen Schätzung von δ ausgehen, welche auf s_1 und s_2 basiert. Für den Fall fehlender Varianzhomogenität kommt folgende Formel zur Anwendung:²⁾

$$\delta_{\text{Diff}} = \sqrt{\frac{\delta_1^2}{N_1} + \frac{\delta_2^2}{N_2}} = \sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}$$

Für unsere empirisch ermittelten Werte ergibt sich

$$\delta_{\text{Diff}} = 0,513$$

$$(M_{\text{B/compr./PU}} = 4,06; M_{\text{A/compr./KU}} = 3,54)$$

$$\boxed{t = 1,014} \quad (df = 58) \quad (t_{\text{krit } 5\%} = 2,005)$$

1) Vgl. 8.12 a)

2) Vgl. E. MITTENECKER, a.a.O., S. 65

3) Aufgrund der nicht vorhandenen Varianzhomogenität beträgt hier die Anzahl der Freiheitsgrade

$$df = \frac{(N_1 + N_2) - 2}{2}$$

(vgl. E. MITTENECKER, a.a.O., S. 65).

H 8: Vergleich A/compr./PU - B/compr./KU:

Da hier die Varianzhomogenität gegeben ist, kann die Verrechnung analog zu 8.13 b) erfolgen.

$$G_{\text{Diff}} = 0,429$$

$$(M_{\text{A/compr./PU}} = 4,17; M_{\text{B/compr./KU}} = 3,12)$$

$$t = 2,448 \quad (df = 117) \quad (t_{\text{krit } 5\%} = 1,984) \\ (t_{\text{krit } 1\%} = 2,626)$$

Zusammenfassung:

H 5: Die Nullhypothese muß verworfen werden

H 6: Die Nullhypothese muß verworfen werden

H 7: Die Nullhypothese muß beibehalten werden

H 8: Die Nullhypothese muß verworfen werden.

8.20 Hypothesen, welche den Zusammenhang von Intelligenz und Lernleistung betreffen

8.21 Normalität der Verteilungen:

Da unsere beiden Gruppen keine Zufallsstichproben sind und sich jede derselben aus drei in sich abgeschlossenen Klasseneinheiten zusammensetzt, halten wir es für notwendig, die Verteilungsform der SW-Werte zu überprüfen. Wir verwendeten dabei das unter Punkt 8.11 geschilderte Verfahren: Dabei ergaben sich folgende Werte¹⁾:

1) Detaillierte Angaben dazu finden sich im Anhang.

$$A_{SW} : \chi^2 = 4,77 \text{ (Chi}_{\text{krit}}^2 5\% = 9,49) \text{ (df} = 4)$$

$$B_{SW} : \chi^2 = 0,80 \text{ (Chi}_{\text{krit}}^2 5\% = 9,49) \text{ (df} = 4)$$

Wir müssen in beiden Fällen die Nullhypothese beibehalten und können somit von einer normalverteilten Grundgesamtheit ausgehen. Die Güte der Anpassung muß bei der Gruppe B als gut, bei der Gruppe A als noch ausreichend bezeichnet werden¹⁾.

8.22 Intelligenz und Lernerfolg in der Lernzielkategorie "knowledge"

Der Zusammenhang der beiden Maßzahlreihen wurde mittels der Formel²⁾

$$r = \frac{N \sum XY - \sum X \sum Y}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}} \quad (\text{df} = N - 2)$$

errechnet. Die unterschiedliche Schiefe der miteinander korrelierten Verteilungen kann zwar als Hinweis auf eine Nichtlinearität des Zusammenhangs gelten. Diese Abweichungen sind jedoch nicht so groß, daß eine signifikante Abweichung vom linearen Zusammenhang befürchtet werden muß, so daß wir nicht mit einer maßgeblichen Verfälschung der Produktmomentkorrelationen rechnen müssen.

1) Vgl. G.A. LIENERT, a.a.O., S. 74

2) Vgl. E. MITTENECKER, a.a.O., S. 104

H 9: Ausprägung des Zusammenhangs von Intelligenz
und Lernleistung:

$$\text{Gruppe A : } r_{A/\text{knowl.}/\text{KU} - \text{SW}} = \underline{0,59} \quad (r_{\text{krit}} 1\% = 0,325)$$

$$r_{A/\text{knowl.}/\text{PU} - \text{SW}} = \underline{0,45} \quad (r_{\text{krit}} 1\% = 0,325)$$

$$\text{Gruppe B : } r_{B/\text{knowl.}/\text{KU} - \text{SW}} = \underline{0,26} \quad (r_{\text{krit}} 5\% = 0,260)$$

$$r_{B/\text{knowl.}/\text{PU} - \text{SW}} = \underline{0,16} \quad (r_{\text{krit}} 5\% = 0,260)$$

Zur Berechnung der durchschnittlichen Korrelationskoeffizienten¹⁾ wurden die Koeffizienten der beiden Gruppen mittels der Fisher'schen z'-Transformation in z-

1) Die Vertrauensgrenzen bei Korrelationskoeffizienten, welche auf einer relativ kleinen Stichprobe beruhen, sind sehr weitgestreckt. Wir haben daher die Koeffizienten der Gruppen A + B zusammengefaßt und werden uns im Rahmen der Interpretation insbesondere auf sie stützen. Die Korrelationskoeffizienten der einzelnen Gruppen sind angesichts des weitgestreckten Vertrauensbereiches im Rahmen der Interpretation nur dann (bedingt) zu verwenden, wenn sie in jeweils signifikanter Ausprägung in derselben Richtung vorliegen. Die numerische Ausprägung der im Rahmen dieser Arbeit angeführten Korrelationskoeffizienten kann auch im Falle der Zusammenfassung unserer beiden Versuchsgruppen nur bedingt interpretiert werden, da zu einer solchermaßen detaillierten Interpretation an und für sich wenigstens 350 Meßwertpaare zugrunde gelegt werden müßten (vgl. E. MITTENECKER, a.a.O., S. 107). Es sei an dieser Stelle vorwegnehmend - wir werden noch im Rahmen der Interpretation darauf zurückkommen - ausdrücklich darauf hingewiesen, daß gerade bei der Interpretation von Korrelationskoeffizienten im Rahmen von Vergleichsstudien sehr häufig unzulässige Schlüsse und darauf basierende Generalisierungen vorgenommen werden.

Werte umgewandelt. Der durchschnittliche z-Wert ergibt sich aus der Formel¹⁾

$$M_z = \frac{\sum_{j=1}^2 (n_j - 3) z'_j}{\sum_{j=1}^2 (n_j - 3)} \quad (df = n_j - 3)$$

Der mittlere Korrelationskoeffizient r_M wurde durch Rücktransformation des solchermaßen errechneten M_z -Wertes ermittelt. Dabei ergaben sich folgende Werte:

$$r_{M_{A+B}/\text{knowl.}/KU} = \boxed{0,45} \quad (r_{\text{krit}} 1\% = 0,234)$$

$$r_{M_{A+B}/\text{knowl.}/PU} = \boxed{0,32} \quad (r_{\text{krit}} 1\% = 0,234)$$

H 10: Unterschiede der Ausprägung des Zusammenhangs zwischen Intelligenz und der Lernleistung in den beiden Unterrichtsmethoden:

a) Zur Feststellung des Stichprobenfehlers der Differenz von Korrelationskoeffizienten verwendeten wir im Falle des Vergleichs zweier unabhängiger Stichproben die Formel²⁾

$$\delta_{z'_1 - z'_2} = \sqrt{\delta_{z'_1}^2 + \delta_{z'_2}^2} = \sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}$$

Mit Hilfe dieses Stichprobenfehlers der Differenz führten wir den z-Test durch, indem wir den kritischen Bruch³⁾

1) Vgl. P.R. HOFSTÄTTER und D. WENDT, a.a.O., S. 162

2) Vgl. E. MITTENECKER, a.a.O., S. 107

3) Vgl. P.R. HOFSTÄTTER und D. WENDT, a.a.O., S. 162

$$z = \frac{z'_1 - z'_2}{\sqrt{\delta_{z_1}^2 + \delta_{z_2}^2}}$$

bildeten (dabei sind z_1 und z_2 die beiden mit Hilfe der Fisher'schen z' -Transformation umgewandelten Korrelationskoeffizienten, deren Differenz von uns untersucht wird). Mittels dieser Rechenoperationen gelangten wir zu folgenden Werten:

Differenz $r_{A/\text{knowl.}/KU} - r_{B/\text{knowl.}/PU} : (0,59 - 0,16)$

$$z = 2,75 \text{ (sign. bei 1 \%)}$$

Differenz¹⁾ $r_{A/\text{knowl.}/KU} - r_{B/\text{knowl.}/KU} : (0,59 - 0,26)$

$$z = 2,19 \text{ (sign. bei 5 \%)}$$

b) Für die Berechnung der Differenz von Korrelationskoeffizienten bei abhängigen Stichproben verwendeten wir den "t-Test für den Unterschied abhängiger Korrelationen":²⁾

$$t = \frac{(r_{12} - r_{13}) \sqrt{(N - 3) (1 + r_{23})}}{\sqrt{2 (1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2 r_{12} r_{13} r_{23})}} \quad (df = N - 3)$$

1) Die hier berechnete Differenz bezieht sich nicht auf unsere Hypothese, sondern wurde lediglich berechnet, um auf die Gefahr der Interpretation von Korrelationskoeffizienten im Rahmen von quasiexperimentellen Anordnungen mit nicht gleichartiger Kontrollgruppe hinzuweisen. Näheres findet sich im Rahmen der Interpretation unserer Versuchsergebnisse.

2) Vgl. E. MITTENECKER, a.a.O., S. 108

Differenz $r_{M_{A+B}/\text{knowl.}/KU} - r_{M_{A+B}/\text{knowl.}/PU} : (0,45 - 0,32)$

$$t = 1,520 \quad (t_{\text{krit } 5\%} = 1,984)$$

Differenz $r_{A/\text{knowl.}/KU} - r_{A/\text{knowl.}/PU} : (0,59 - 0,45)$

$$t = 1,259 \quad (t_{\text{krit } 5\%} = 2,000)$$

H 11: Unterschiede der Lehrwirksamkeit von Programmierter Unterweisung und herkömmlichem Unterricht in Extremgruppen:

Zur Bildung der Extremgruppen zogen wir die Versuchspersonen im oberen bzw. unteren Fünftel unserer empirisch gefundenen Intelligenzverteilungen heran.

Gruppe A	Gruppe B
SW $\searrow$ 125 : $N_2 = 15$	SW $\searrow$ 124 : $N_1 = 12$
SW $\swarrow$ 109 : $N_1 = 11$	SW $\swarrow$ 108 : $N_2 = 13$

Da bei Verwendung von Extremgruppen nicht mehr von parametrischen Bedingungen ausgegangen werden kann, verwendeten wir zum Vergleich der Maßzahldifferenzen (PU - KU) den Wilcoxon-Whyte-Test¹⁾. Mit Hilfe dieses Verfahrens können wir feststellen, ob die Rangteilsomme T_1 der kleineren Gruppe von dem theoretisch zu erwartenden Wert T_{pop} überzufällig abweicht. Dabei bedeuten:

T_1 = Tatsächlich beobachtete Summe aller Ränge der kleineren Gruppe

$T_{\text{pop}} = \frac{N_1 (N + 1)}{2}$, d.h. theoretisch erwartete Summe aller Ränge der kleineren Gruppe

1) Vgl. E. MITTENECKER, a.a.O., S. 161 ff.

N_1 = Anzahl der Versuchspersonen der kleineren Gruppe

N = Anzahl der Versuchspersonen beider Extremgruppen.

Als Prüfgröße dient der Ausdruck¹⁾

$$|T_1 - T_{pop}|$$

Mittels der oben angegebenen Formeln errechneten wir für die beiden Gruppen A und B folgende Werte:

$$T_1 - T_{pop} = - 9,5$$

nicht sign.

$$T_1 - T_{pop} = - 19$$

nicht sign.

Zusammenfassung:

- H 9: Für die Gesamtgruppe A+B ergab sich bei beiden Unterrichtsmethoden ein signifikanter Zusammenhang von Lernleistung und Intelligenz auf dem 1%-Niveau. Bei der Gruppe A haben wir in beiden Fällen einen signifikanten Zusammenhang auf dem 1%-Niveau, bei der Gruppe B fand sich lediglich im frontalunterrichtlichen Verfahren ein signifikanter Zusammenhang auf dem 5%-Niveau.
- H 10: Die Nullhypothese kann bei Zugrundelegung der Gesamtgruppe A+B nicht verworfen werden, d.h. es ist kein signifikanter Unterschied in der von uns erwarteten Richtung feststellbar. Es kann allerdings von einer Tendenz in Richtung unserer Hypothese gesprochen werden.
- H 11: Die Nullhypothese muß beibehalten werden, d.h. es konnten keine signifikanten Unterschiede der Lehrwirksamkeit von Programmierter Unterweisung und herkömmlichem Unterricht bei den beiden Extremgruppen festgestellt werden.

1) Vgl. E. MITTENECKER, a.a.O., S. 164

8.23 Intelligenz und Lernerfolg in der Lernzielkategorie "comprehension"

Die Verrechnung der Daten erfolgte analog zu Punkt 8.22.
Dabei ergaben sich folgende Ergebnisse:

H 12: Ausprägung des Zusammenhangs von Intelligenz und Lernleistung:

$$\text{Gruppe A : } r_{A/\text{compr.}/\text{KU-SW}} = \underline{0,29} \quad (r_{\text{krit}} 5\% = 0,25)$$

$$r_{A/\text{compr.}/\text{PU-SW}} = \underline{0,19} \quad (r_{\text{krit}} 5\% = 0,25)$$

$$\text{Gruppe B : } r_{B/\text{compr.}/\text{KU-SW}} = \underline{0,17} \quad (r_{\text{krit}} 5\% = 0,26)$$

$$r_{B/\text{compr.}/\text{PU-SW}} = \underline{0,33} \quad (r_{\text{krit}} 5\% = 0,26)$$

$$(r_{\text{krit}} 1\% = 0,34)$$

Mittlere Korrelationskoeffizienten (A+B):

$$r_{M_{A+B}/\text{compr.}/\text{KU-SW}} = \underline{0,24} \quad (r_{\text{krit}} 1\% = 0,234)$$

$$r_{M_{A+B}/\text{compr.}/\text{PU-SW}} = \underline{0,25} \quad (r_{\text{krit}} 1\% = 0,234)$$

H 13: Unterschiede der Ausprägung des Zusammenhangs zwischen Intelligenz und der Lernleistung in den beiden Unterrichtsmethoden:

$$r_{A/\text{compr.}/\text{KU-SW}} - r_{B/\text{compr.}/\text{PU-SW}} : 0,29 - (+0,33) = \underline{-0,04}$$

$$r_{B/\text{compr.}/\text{KU-SW}} - r_{A/\text{compr.}/\text{PU-SW}} : 0,17 - (+0,19) = \underline{-0,02}$$

$$r_{M_{A+B}/\text{compr.}/\text{KU-SW}} - r_{M_{A+B}/\text{compr.}/\text{PU-SW}} : 0,24 - (+0,25) = \underline{-0,01}$$

$$r_{A/\text{compr.}/\text{KU-SW}} - r_{A/\text{compr.}/\text{PU-SW}} : 0,29 - (+0,19) = \underline{0,10}$$

$$r_{B/\text{compr.}/\text{KU-SW}} - r_{B/\text{compr.}/\text{PU-SW}} : 0,17 - (+0,33) = \underline{0,16}$$

Sämtliche Unterschiede sind nichtsignifikant.

H 14: Unterschiede der Lehrwirksamkeit von Programmierter Unterweisung und herkömmlichem Unterricht in Extremgruppen:

Die Gruppeneinteilung entspricht der unter H 11 aufgeführten.

Gruppe A	Gruppe B
$T_1 - T_{\text{pop}} = -4,0$	$T_1 - T_{\text{pop}} = -34$
nicht sign.	nicht sign.
	$ T_1 - T_{\text{pop}} _{\text{krit } 5\%} = 35$

Zusammenfassung:

H 12: Für die Gesamtgruppe A+B ergab sich bei beiden Unterrichtsmethoden ein signifikanter Zusammenhang auf dem 1%-Niveau. Die Koeffizienten der Gruppe A bzw. B sind jeweils nur einmal abgesichert, und zwar bei der Gruppe A auf dem 5%-Niveau, bei der Gruppe B knapp unter dem 1%-Niveau.

H 13: Die Nullhypothese kann in keinem Fall verworfen werden. Es ist auch keine Tendenz in Richtung unserer Hypothese erkennbar.

H 14: Die Nullhypothese muß beibehalten werden, d.h. es konnten keine signifikanten Unterschiede der Lehrwirksamkeit von Programmierter Unterweisung und herkömmlichem Unterricht bei den beiden Extremgruppen festgestellt werden.

8.30 Hypothesen, welche die "positiven Verstärkungen" betreffen

8.31 Normalität der Verteilungen:¹⁾

A_{Verst.} : $\chi^2 = 1,38$ ($\chi^2_{\text{krit}} 5\% = 9,49$) (df = 4)

B_{Verst.} : $\chi^2 = 1,49$ ($\chi^2_{\text{krit}} 5\% = 9,49$) (df = 4)

Die Nullhypothese kann nicht verworfen werden, die Anpassung an die Normalverteilung kann als gut bezeichnet werden. Die Voraussetzungen zur Anwendung der Produkt-Moment-Korrelation sind erfüllt²⁾.

8.32 Das Ausmaß an positiven Verstärkungen im herkömmlichen Unterricht und die Lernleistung in den Lernzielkategorien "knowledge" und "comprehension":

H 15: Ausprägung des Zusammenhangs³⁾ zwischen der Lernleistung in den beiden Lernzielkategorien im herkömmlichen Unterricht und der Anzahl an positiven Verstärkungen:

-
- 1) Detaillierte Angaben finden sich im Anhang.
 - 2) Bezüglich der Linearität der untersuchten Zusammenhänge gehen wir von den unter Punkt 8.22 festgestellten Bedingungen aus. Dies gilt auch für die weiteren Korrelationsberechnungen.
 - 3) Eine partielle Korrelation unter Ausschluß der gemeinsamen Intelligenzkomponente wurde nicht durchgeführt. Die qualitative Analyse der fraglichen Kategorien ergab hierzu keinen zwingenden Hinweis. Dies wurde auch durch eine Überprüfung des Zusammenhangs der Häufigkeit an positiven Verstärkungen und dem SW-Wert bestätigt.
- $r_{A/\text{Verst.}}\text{-SW}=0,01$; $r_{B/\text{Verst.}}\text{-SW}=0,07$; $r_{M_{A+B}/\text{Verst.}}\text{-SW}=0,04$

Die Verrechnung der Daten erfolgte analog zu oben

Gruppe A : $r_{A/\text{Verst.}-\text{knowl.}/\text{KU}} = \underline{0,28}$ ($r_{\text{krit}} 5\% = 0,250$)

$r_{A/\text{Verst.}-\text{compr.}/\text{KU}} = \underline{0,17}$ ($r_{\text{krit}} 1\% = 0,325$)

Gruppe B : $r_{B/\text{Verst.}-\text{knowl.}/\text{KU}} = \underline{-0,32}$ ($r_{\text{krit}} 1\% = 0,340$)

$r_{B/\text{Verst.}-\text{compr.}/\text{KU}} = \underline{-0,16}$ ($r_{\text{krit}} 5\% = 0,260$)

Mittlere Korrelationskoeffizienten (A+B):

$r_{M_{A+B}/\text{Verst.}-\text{knowl.}/\text{KU}} = \underline{0,05}$

$r_{M_{A+B}/\text{Verst.}-\text{compr.}/\text{KU}} = \underline{-0,003}$ ($r_{\text{krit}} 5\% = 0,183$)

H 16: Unterschied der Lehrwirksamkeit von Programmierter Unterweisung und herkömmlichem Unterricht in Extremgruppen:

Die Gruppeneinteilung erfolgte analog zu Punkt 8.22 .

Gruppe A

Gruppe B¹⁾

Verstärk.: ♣ 15 : N₂ = 11

Verstärk.: ♣ 24 : N₁ = 8

Verstärk.: ♠ 6 : N₁ = 10

Verstärk.: ♠ 11 : N₂ = 11

knowl.

compr.

Gruppe A	$T_1 - T_{\text{pop}} = -12$ nicht sign.	$T_1 - T_{\text{pop}} = -6,5$ nicht sign.
Gruppe B	$T_1 - T_{\text{pop}} = -26,5$ sign. bei 5%	$T_1 - T_{\text{pop}} = +16$ nicht sign.

1) Der Zusammenhang zwischen Intelligenz und der Häufigkeit an positiven Verstärkungen ist so gering, daß keine systematischen Intelligenzunterschiede zwischen den Gruppen aus dem oberen und unteren Bereich der Verteilung zu erwarten sind.

Zusammenfassung:

- H 15: Die Nullhypothese muß beibehalten werden, d.h. daß kein signifikanter Zusammenhang zwischen der Anzahl an positiven Verstärkungen und der Lernleistung im herkömmlichen Unterricht festgestellt werden kann.
- H 16: Bei der Gruppe A muß die Nullhypothese für beide Lernzielkategorien beibehalten werden. Bei der Gruppe B ist in der Lernzielkategorie "knowledge" der mittlere Rang der Differenzen $X_{PU} - X_{KU}$ bei denjenigen Schülern, welche im herkömmlichen Unterricht extrem häufig verstärkt wurden, signifikant höher (d.h. numerisch geringer).

8.40 Hypothesen, welche die "unterrichtsbezogene Schüleraktivität" betreffen

8.41 Normalität der Verteilungen:¹⁾

A Antw.: $\chi^2 = 5,42$ ($\chi^2_{krit\ 5\%} = 9,49$) (df = 4)

B Antw.: $\chi^2 = 4,31$ ($\chi^2_{krit\ 5\%} = 9,49$) (df = 4)

Die Nullhypothese kann nicht verworfen werden, die Güte der Anpassung an die Normalverteilung ist mäßig. Die Voraussetzungen zur Anwendung der Produkt-Moment-Korrelation sind erfüllt.²⁾

1) Detaillierte Angaben finden sich im Anhang.

2) Wir verstehen den Begriff "unterrichtsbezogene Schüleraktivität" ausschließlich als die Frequenz aller einschlägigen Beobachtungskategorien. Bezüglich des Skalierungsniveaus sind damit Bedingungen gegeben, wie sie z.B. bei der Lidschlagfrequenz vorliegen (vgl. E. MITTENECKER, a.a.O., S. 121).

8.42 Das Ausmaß an "unterrichtsbezogener Schüleraktivität" im herkömmlichen Unterricht und die Lernleistung in den Lernzielkategorien "knowledge" und "comprehension"

H 17: Ausprägung des Zusammenhangs¹⁾ zwischen der Lernleistung im herkömmlichen Unterricht und dem Ausmaß an "stoffbezogener Schüleraktivität":

Gruppe A: $r_{A/\text{Antw.}/\text{knowl.}/\text{KU}} = \underline{0,24}$ ($r_{\text{krit } 5\%} = 0,250$)

$r_{A/\text{Antw.}/\text{compr.}/\text{KU}} = \underline{0,15}$ ($r_{\text{krit } 5\%} = 0,250$)

Gruppe B: $r_{B/\text{Antw.}/\text{knowl.}/\text{KU}} = \underline{-0,15}$ ($r_{\text{krit } 5\%} = 0,260$)

$r_{B/\text{Antw.}/\text{compr.}/\text{KU}} = \underline{-0,17}$ ($r_{\text{krit } 5\%} = 0,260$)

Mittlere Korrelationskoeffizienten (A+B):

$r_{M_{A+B}/\text{Antw.}/\text{knowl.}/\text{KU}} = \underline{0,05}$ ($r_{\text{krit } 5\%} = 0,183$)

$r_{M_{A+B}/\text{Antw.}/\text{compr.}/\text{KU}} = \underline{-0,01}$ ($r_{\text{krit } 5\%} = 0,183$)

H 18: Unterschied der Lehrwirksamkeit von Programmierter Unterweisung und herkömmlichem Unterricht in Extremgruppen:

Die Gruppenteilung erfolgte analog zu oben.

1) Auch hier wurde keine partielle Korrelation unter Ausschluß der gemeinsamen Intelligenzkomponente durchgeführt. Die Überprüfung des Zusammenhangs der Häufigkeit an unterrichtsbezogener Schüleraktivität und dem SW-Wert ergab folgende Koeffizienten:

$r_{A/\text{Akt.}-\text{SW}} = 0,22$; ($r_{\text{krit } 5\%} = 0,25$);

$r_{B/\text{Akt.}-\text{SW}} = 0,13$; $r_{M_{A+B}/\text{Akt.}-\text{SW}} = 0,15$ ($r_{\text{krit } 5\%} = 0,183$)

Gruppe A

Gruppe B¹⁾

unterrichtsb.Akt. $\gg 21:N_1 = 8$

unterrichtsb.Akt. $\gg 31:N_1 = 9$

unterrichtsb.Akt. $\ll 9:N_2 = 11$

unterrichtsb.Akt. $\ll 10:N_2 = 11$

knowl.

compr.

Gruppe A

$T_1 - T_{pop} = -6$
nicht sign.

$T_1 - T_{pop} = -5$
nicht sign.

Gruppe B

$T_1 - T_{pop} = -29,5$
sign. bei 5%

$T_1 - T_{pop} = +7,5$
nicht sign.

Zusammenfassung:

H 17: Die Nullhypothese muß beibehalten werden, d.h. es kann kein signifikanter Zusammenhang zwischen dem Ausmaß an "unterrichtsbezogener Schüleraktivität" und der Lernleistung im herkömmlichen Unterricht festgestellt werden.

H 18: Bei der Gruppe A muß die Nullhypothese für beide Lernzielkategorien beibehalten werden. Bei der Gruppe B ist in der Lernzielkategorie "knowledge" der mittlere Rang der Differenzen $X_{PU} - X_{KU}$ bei denjenigen Schülern, welche im herkömmlichen Unterricht extrem häufig "unterrichtsbezogene Aktivität" zeigten, signifikant höher (d.h. numerisch geringer).

1) Aufgrund der nichtsignifikanten Zusammenhänge zwischen Intelligenz und dem Ausmaß an "unterrichtsbezogener Schüleraktivität" sind keine systematischen Intelligenzunterschiede zwischen den Gruppen aus dem oberen und unteren Bereich der "Aktivitätsverteilung" zu erwarten.

8.50 Hypothesen, welche die "konzentrativen Arbeitshaltung" betreffen

8.51 Normalität der Verteilungen:¹⁾

$$^A \text{Arb.konz. } \chi^2 = 3,65 \text{ (Chi}_{\text{krit}}^2 5\% = 9,49) \text{ (df} = 4)$$

$$^B \text{Arb.konz. } \chi^2 = 3,45 \text{ (Chi}_{\text{krit}}^2 5\% = 9,49) \text{ (df} = 4)$$

Die Nullhypothese kann nicht verworfen werden, die Güte der Anpassung an die Normalverteilung ist mäßig.

8.52 "Konzentrativen Arbeitshaltung" und Lernleistung in den Lernzielkategorien "knowledge" und "comprehension":

H 19: Die Ausprägung des Zusammenhangs zwischen der Lernleistung in den beiden Lernzielkategorien und der "konzentrativen Arbeitshaltung:

$$\text{Gruppe A : } r_{A/\text{knowl.}/\text{KU-Arb.konz.}} = \underline{0,48} \text{ (} r_{\text{krit}} 1\% = 0,325)$$

$$r_{A/\text{knowl.}/\text{PU-Arb.konz.}} = \underline{0,38} \text{ (} r_{\text{krit}} 1\% = 0,325)$$

$$r_{A/\text{compr.}/\text{KU-Arb.konz.}} = \underline{0,53} \text{ (} r_{\text{krit}} 1\% = 0,325)$$

$$r_{A/\text{compr.}/\text{PU-Arb.konz.}} = \underline{0,59} \text{ (} r_{\text{krit}} 1\% = 0,325)$$

$$\text{Gruppe B : } r_{B/\text{knowl.}/\text{KU-Arb.konz.}} = \underline{0,43} \text{ (} r_{\text{krit}} 1\% = 0,340)$$

$$r_{B/\text{knowl.}/\text{PU-Arb.konz.}} = \underline{0,27} \text{ (} r_{\text{krit}} 5\% = 0,260)$$

$$r_{B/\text{compr.}/\text{KU-Arb.konz.}} = \underline{0,26} \text{ (} r_{\text{krit}} 5\% = 0,260)$$

$$r_{B/\text{compr.}/\text{PU-Arb.konz.}} = \underline{0,24} \text{ (} r_{\text{krit}} 5\% = 0,260)$$

1) Detaillierte Angaben finden sich im Anhang.

Für die Berechnung des tatsächlichen Zusammenhangs zwischen der "konzentrativen Arbeitshaltung" und der Lernleistung wurde eine partielle Korrelation durchgeführt, da die Korrelation dieser beiden Variablen dadurch verfälscht wird, daß beide ihrerseits mit der Intelligenz relativ hoch korrelieren. Dies folgt sowohl aus der qualitativen Analyse der Ratingskalen zur Ermittlung der "konzentrativen Arbeitshaltung" als auch aufgrund der quantitativen Analyse des Zusammenhangs der Variablen "Intelligenz" und "Lernleistung" auf der einen und des Zusammenhangs von "Intelligenz" und "konzentrativer Arbeitshaltung" auf der anderen Seite. Dabei ergaben sich für den Zusammenhang von Intelligenz und "konzentrativer Arbeitshaltung" folgende Werte:¹⁾

$$\begin{aligned} r_{A/SW-Arb.konz.} &= 0,49 \\ r_{B/SW-Arb.konz.} &= 0,46 \end{aligned} \quad \begin{array}{l} \text{(beide Koeffizienten sind auf} \\ \text{dem 1\%-Niveau signifikant)} \end{array}$$

Zur Berechnung der partiellen Korrelation wurde folgende Formel verwendet:²⁾

$$r_{12-3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{(1 - r_{13}^2) (1 - r_{23}^2)}}$$

Dabei ergaben sich folgende Werte:

$$\text{Gruppe A: } r_{A_{part.}/knowl./KU} = \underline{0,38} \quad (r_{krit} \ 1\% = 0,325)$$

$$r_{A_{part.}/knowl./PU} = \underline{0,20} \quad (r_{krit} \ 5\% = 0,250)$$

$$r_{A_{part.}/compr./KU} = \underline{0,47} \quad (r_{krit} \ 1\% = 0,325)$$

$$r_{A_{part.}/compr./PU} = \underline{0,58} \quad (r_{krit} \ 1\% = 0,325)$$

1) Die Koeffizienten für den Zusammenhang von Intelligenz und Lernleistung sind unter Punkt 8.20 aufgeführt.

2) Vgl. E. MITTENECKER, a.a.O., S. 110

Gruppe B: $r_{B_{part.}/knowl./KU} = \underline{0,36}$ ($r_{krit} 1\% = 0,340$)

$r_{B_{part.}/knowl./PU} = \underline{0,22}$ ($r_{krit} 5\% = 0,260$)

$r_{B_{part.}/compr./KU} = \underline{0,22}$ ($r_{krit} 5\% = 0,260$)

$r_{B_{part.}/compr./PU} = \underline{0,11}$ ($r_{krit} 5\% = 0,260$)

Die Berechnung der mittleren Korrelationskoeffizienten der Gruppen A+B auf der Basis der partiellen Korrelation ergab folgende Werte:

$r_{M_{A+B_{part.}}/knowl./KU} = \underline{0,37}$ ($r_{krit} 1\% = 0,234$)

$r_{M_{A+B_{part.}}/knowl./PU} = \underline{0,21}$ ($r_{krit} 5\% = 0,183$)

$r_{M_{A+B_{part.}}/compr./KU} = \underline{0,36}$ ($r_{krit} 1\% = 0,234$)

$r_{M_{A+B_{part.}}/compr./PU} = \underline{0,40}$ ($r_{krit} 1\% = 0,234$)

H 20: Unterschied des Zusammenhangs zwischen Lernleistung und "konzentrativer Arbeitshaltung" bei den beiden Unterrichtsmethoden:

Differenz $r_{M_{A+B}/knowl./KU} - r_{M_{A+B}/knowl./PU}$: (0,37-0,21)

$$t = \underline{1,844} \quad (t_{krit} 5\% = 1,984)$$

Alle übrigen Differenzen sind ebenfalls nichtsignifikant. Die Differenzen $r_{A_{part.}}/knowl./KU - r_{A_{part.}}/knowl./PU$ (0,38 - 0,20) und $r_{B_{part.}}/knowl./KU - r_{B_{part.}}/knowl./PU$ (0,36 - 0,22) weisen allerdings deutlich in Richtung unserer Hypothese.

Zusammenfassung:

- H 19: Die Nullhypothese muß bei Zugrundelegung der mittleren Korrelationskoeffizienten der Gruppen A+B grundsätzlich verworfen werden, d.h. es besteht bei beiden Lernzielkategorien in beiden Unterrichtsmethoden ein signifikanter Zusammenhang zwischen "konzentrativer Arbeitshaltung" und der Lernleistung.
- H 20: Die Nullhypothese muß bei allen Differenzen beibehalten werden, allerdings zeigt sich im Bereich der Lernzielkategorie "knowledge" eine deutliche Tendenz in Richtung unserer Hypothese.

8.60 "Einstellung zur Schule

8.61 Normalität der Verteilungen:¹⁾

$$A_{\text{Einst.z.Sch.}} \chi^2 = 2,68 \text{ (Chi}_{\text{krit}}^2 = 9,49) \text{ (df} = 4)$$

$$B_{\text{Einst.z.Sch.}} \chi^2 = 2,73 \text{ (Chi}_{\text{krit}}^2 = 9,49) \text{ (df} = 4)$$

Die Nullhypothese kann nicht verworfen werden, die Güte der Anpassung an die Normalverteilung ist ausreichend.

8.62 Die "Einstellung zur Schule" und die Lernleistung in den beiden Lernzielkategorien:

H 21: Unterschiede der Ausprägung des Zusammenhangs zwischen der Einstellung zur Schule und der Lernlei-

1) Nähere Angaben finden sich im Anhang.

stung in den beiden Unterrichtsmethoden:

$$\text{Gruppe A: } r_{A/\text{knowl.}/\text{KU-Einst.}} = \underline{-0,02} \quad (r_{\text{krit}} 5\% = 0,250)$$

$$r_{A/\text{knowl.}/\text{PU-Einst.}} = \underline{-0,002} \quad (r_{\text{krit}} 5\% = 0,250)$$

$$r_{A/\text{compr.}/\text{KU-Einst.}} = \underline{0,28} \quad (r_{\text{krit}} 5\% = 0,250)$$

$$r_{A/\text{compr.}/\text{PU-Einst.}} = \underline{-0,04} \quad (r_{\text{krit}} 5\% = 0,250)$$

$$\text{Gruppe B: } r_{B/\text{knowl.}/\text{KU-Einst.}} = \underline{0,06} \quad (r_{\text{krit}} 5\% = 0,260)$$

$$r_{B/\text{knowl.}/\text{PU-Einst.}} = \underline{-0,26} \quad (r_{\text{krit}} 5\% = 0,260)$$

$$r_{B/\text{compr.}/\text{KU-Einst.}} = \underline{-0,10} \quad (r_{\text{krit}} 5\% = 0,260)$$

$$r_{B/\text{compr.}/\text{PU-Einst.}} = \underline{0,03} \quad (r_{\text{krit}} 5\% = 0,260)$$

Mittlere Korrelationskoeffizienten (A+B):

$$r_{M_{A+B}/\text{knowl.}/\text{KU-Einst.}} = \underline{0,02} \quad (r_{\text{krit}} 5\% = 0,183)$$

$$r_{M_{A+B}/\text{knowl.}/\text{PU-Einst.}} = \underline{-0,13} \quad (r_{\text{krit}} 5\% = 0,183)$$

$$r_{M_{A+B}/\text{compr.}/\text{KU-Einst.}} = \underline{0,10} \quad (r_{\text{krit}} 5\% = 0,183)$$

$$r_{M_{A+B}/\text{compr.}/\text{PU-Einst.}} = \underline{0,007} \quad (r_{\text{krit}} 5\% = 0,183)$$

Zusammenfassung:

H 21: Die mittleren Korrelationskoeffizienten sind alle nichtsignifikant. Dies gilt auch für deren Differenzen. Wir müssen die Nullhypothese also beibehalten.

8.70 Einstellung zum Lehrer

8.71 Normalität der Verteilungen:¹⁾

$$A_{\text{Einst.z.L.}} \chi^2 = 3,98 \quad (\chi^2_{\text{krit}} = 9,49) \quad (df = 4)$$

$$B_{\text{Einst.z.L.}} \chi^2 = 5,19 \quad (\chi^2_{\text{krit}} = 9,49) \quad (df = 4)$$

Die Nullhypothese kann nicht verworfen werden, die Güte der Anpassung an die Normalverteilung ist mäßig.

8.72 Die Einstellung zum Lehrer und die Lernleistung in den beiden Lernzielkategorien:

H 22: Unterschiede der Ausprägung des Zusammenhangs zwischen der Einstellung zum Lehrer und der Lernleistung in den beiden Unterrichtsmethoden:

Gruppe A: $r_{A/\text{knowl.}/\text{KU-Einst.z.L.}} = \underline{0,08} \quad (r_{\text{krit } 5\%} = 0,250)$

$$r_{A/\text{knowl.}/\text{PU-Einst.z.L.}} = \underline{0,07} \quad (r_{\text{krit } 5\%} = 0,250)$$

$$r_{A/\text{compr.}/\text{KU-Einst.z.L.}} = \underline{-0,01} \quad (r_{\text{krit } 5\%} = 0,250)$$

$$r_{A/\text{compr.}/\text{PU-Einst.z.L.}} = \underline{-0,26} \quad (r_{\text{krit } 5\%} = 0,250)$$

Gruppe B: $r_{B/\text{knowl.}/\text{KU-Einst.z.L.}} = \underline{0,07} \quad (r_{\text{krit } 5\%} = 0,260)$

$$r_{B/\text{knowl.}/\text{PU-Einst.z.L.}} = \underline{0,13} \quad (r_{\text{krit } 5\%} = 0,260)$$

$$r_{B/\text{compr.}/\text{KU-Einst.z.L.}} = \underline{0,09} \quad (r_{\text{krit } 5\%} = 0,260)$$

$$r_{B/\text{compr.}/\text{PU-Einst.z.L.}} = \underline{0,25} \quad (r_{\text{krit } 5\%} = 0,260)$$

1) Nähere Angaben finden sich im Anhang.

Mittlere Korrelationskoeffizienten (A+B):

$$r_{M_{A+B}/\text{knowl.}/\text{KU-Einst.z.L.}} = \underline{0,08} \quad (r_{\text{krit } 5\%} = 0,183)$$

$$r_{M_{A+B}/\text{knowl.}/\text{PU-Einst.z.L.}} = \underline{0,10} \quad (r_{\text{krit } 5\%} = 0,183)$$

$$r_{M_{A+B}/\text{compr.}/\text{KU-Einst.z.L.}} = \underline{0,04} \quad (r_{\text{krit } 5\%} = 0,183)$$

$$r_{M_{A+B}/\text{compr.}/\text{PU-Einst.z.L.}} = \underline{-0,02} \quad (r_{\text{krit } 5\%} = 0,183)$$

Sämtliche Differenzen zwischen den Korrelationskoeffizienten sind nichtsignifikant.

Zusammenfassung:

H 22: Die Nullhypothese muß sowohl bei den Einzelgruppen als auch bei der Zusammenfassung der Gruppen A+B beibehalten werden. Die mittleren Korrelationskoeffizienten sind alle nichtsignifikant.

9.00 Diskussion der Versuchsergebnisse

Im Zusammenhang mit der Interpretation der Daten der statistischen Analyse sind folgende Gesichtspunkte zu beachten:

- Die Interpretation kann nur auf der Basis der verwendeten Lehrprogramme, unseres herkömmlichen Vergleichsunterrichts und der untersuchten Adressatengruppe (Gymnasium, 10. Schülerjahrgang) erfolgen, weitergehende Generalisierungen sind nicht oder nur in äußerst begrenztem Umfang möglich;
- Angesichts der Tatsache, daß feldexperimentelle Untersuchungen bei Rückgriff auf natürliche Klasseneinheiten i.allg. in weit größerem Ausmaß als "Laboruntersuchungen" systematische Fehlervarianz erwarten lassen, werden wir bei der Interpretation streng darauf zu achten haben, daß nur dort Effekte der von uns gesetzten Bedingungen registriert und interpretiert werden, wo dieselben wenigstens doppelt abgesichert sind;
- Bei der Interpretation der Korrelationskoeffizienten werden wir uns angesichts des weitgestreckten Vertrauensbereichs bei kleinen Stichproben insbesondere auf die mittleren Koeffizienten der Gruppen A+B stützen. Bei den Einzelgruppen A und B werden wir Korrelationskoeffizienten und deren Differenzen nur dann interpretieren, wenn beide in dieselbe Richtung weisen.

9.10 Genereller Lernleistungsvergleich

9.11 Lernzielkategorie "knowledge" (H 1-4)

Die Schüler der Gruppe B weisen bei Programmierter Unterweisung einen signifikant höheren Lernerfolg auf als im herkömmlichen Vergleichsunterricht. Derselbe Befund zeigt sich bei der Gruppe A. Beim Vergleich unabhängiger Stichproben zeigt die Gruppe B im programmierten Unterricht eine signifikant bessere Lernleistung als die herkömmlich unterrichtete Gruppe A, umgekehrt, also beim Vergleich der programmiert unterrichteten Gruppe A mit der herkömmlich unterrichteten Gruppe B, ist jedoch der Unterschied nichtsignifikant. Bezüglich des nichtsignifikanten Befunds muß einschränkend darauf hingewiesen werden, daß einmal der bei unabhängigen Stichproben deutlich größere Stichprobenfehler der Differenz der Mittelwerte mögliche Methodeneffekte eher verdeckt, als dies beim Vergleich abhängiger Stichproben der Fall ist, daß zum anderen die Differenz der arithmetischen Mittel in die von uns erwartete Richtung weist und daß schließlich bereits bei der Analyse möglicher Übungseffekte festgestellt wurde, daß bei der hier diskutierten Vergleichsrichtung durch einen nichtlinearen Übungsgewinn die Ergebnisse gegen die Richtung unserer Hypothese verschoben werden. Wir glauben daher angesichts der dreifachen signifikanten Absicherung und des immerhin trotz ungünstiger Bedingungen in Richtung unserer Hypothese weisenden nichtsignifikanten Unterschieds feststellen zu können, daß die Programmierte Unterweisung sich zum Erlernen einfacher Rechentechniken, wie sie im Rahmen der Lernzielkategorie "knowledge" geprüft worden sind, eindeutig besser eignet als unsere Form des herkömmlichen Vergleichsunterrichts.

Dieses Ergebnis scheint uns umso gewichtiger zu sein, als im Gegensatz zu den meisten einschlägigen Untersuchungen die Programmgruppe nicht durch eine programmorientierte Itemauswahl von vornherein begünstigt wurde.

9.12 Lernzielkategorie "comprehension" (H 5-8)

Hier zeigt sich, daß sowohl die Schüler der Gruppe A als auch der Gruppe B die Testitems in signifikant höherem Ausmaß lösten, wenn sie programmiert unterrichtet wurden. Beim gegenseitigen Vergleich der Gruppen (unabhängige Stichproben) zeigt sich nur einmal ein signifikanter Unterschied der Lernleistung zugunsten der programmiert unterrichteten Gruppe, im zweiten Fall liegt zwar ein Mittelwertsunterschied in Richtung unserer Hypothese vor, dieser ist jedoch nichtsignifikant. Da auch hier mit einer Verdeckung der Methodeneffekte durch den bei unabhängigen Stichproben relativ großen Stichprobenfehler der Mittelwertsdifferenz gerechnet werden muß und darüber hinaus das Ergebnis in Richtung unserer Hypothese weist, können wir zusammenfassend feststellen, daß auch bei der Lernzielkategorie "comprehension" sich die Programmierte Unterweisung als die Methode mit der eindeutig höheren Lehreffektivität erwiesen hat. Dieser Befund scheint uns besonders bedeutsam zu sein, da diejenigen Autoren, welche der Programmierten Unterweisung ablehnend oder zurückhaltend gegenüberstehen, häufig darauf hinweisen, daß die Programmierte Unterweisung zwar "mechanische Kenntnisse und Fertigkeiten" vermitteln könne, selbständige Denkleistung jedoch nicht fördere, ja sogar unterdrücke. Wir können zwar nicht in Anspruch nehmen, mit unserer Lernzielkategorie "comprehension" die Fähigkeit zu "schöpfe-

rischem Denken" vollständig erfaßt zu haben, gehen aber davon aus, daß dies bezüglich einiger operationalisierter Aspekte gelungen ist. Die signifikant höhere Lehr-effektivität der Programmierten Unterweisung in beiden Lernzielkategorien scheint uns umso bedeutsamer zu sein, als unsere Einstellungsmessungen, wie wir schon an anderer Stelle berichteten, eine jeweils hochsignifikante Bevorzugung des Lehrerunterrichts erbrachten und zudem die von uns gewählte Form des Lehrerunterrichts in der einschlägigen didaktischen Literatur als die zur Zeit wirksamste frontalunterrichtliche Methode dargestellt wird.¹⁾

-
- 1) Die durchschnittliche, pro Minute geäußerte Häufigkeit an unterrichtsbezogener Lehreraktivität betrug in den Klassen der Gruppe A 8,7; 10,7; 8,4; bei der Gruppe B lauten die entsprechenden Werte 10,6; 9,7 und 11,7. Demgegenüber ergaben sich für die pro Minute durchschnittlich geäußerte unterrichtsbezogene Schüleraktivität die Werte 1,4; 2,1; 1,2 (Gruppe A) und 2,1; 0,9; 2,9 (Gruppe B). An dieser Stelle sei noch einmal darauf hingewiesen, daß sich diese Werte ausschließlich auf verbal geäußerte Verhaltensweisen beziehen. Dies ist vor allem für die Interpretation des "tatsächlichen Ausmaßes" der Schüleraktivität von Bedeutung. Die gute Disziplinlage der Versuchsklassen kann aus der absoluten Häufigkeit an disziplinären Verlaufsäußerungen entnommen werden (19; 2; 17 (Gruppe A) und 29; 15; 2 (Gruppe B)). Diese Werte beziehen sich auf die gesamten fünf Unterrichtsstunden. Diese geringe Anzahl an disziplinären Verlaufsäußerungen gewinnt besondere Bedeutung, wenn gleichzeitig die durchgehende signifikante Bevorzugung des Lehrerunterrichts mit in Betracht gezogen wird. Wir können von hier aus auf eine entspannte, demokratische und stoffzentrierte Unterrichtslage schließen. Dieser Befund wurde durch die qualitative Beurteilung der Unterrichtslage durch die Beobachter bestätigt.

9.20 Zusammenhang von Intelligenz und Lernleistung

9.21 Lernzielkategorie "knowledge" (H 9-11)

- a) Der mittlere Zusammenhang (Gruppe A+B) ist bei beiden Unterrichtsmethoden signifikant. Bei den Einzelgruppen ergab sich lediglich bei der Gruppe B im Falle der Programmierten Unterweisung ein nichtsignifikanter Zusammenhang. Dieser fällt angesichts der zugrunde liegenden Stichprobengröße und des damit weitgestreckten Vertrauensbereiches nicht ins Gewicht. Es kann daher festgestellt werden, daß auch bei Programmierter Unterweisung ein signifikanter Zusammenhang zwischen der Intelligenz und der Lernleistung besteht. Eine von der Intelligenz relativ unabhängige Manipulierbarkeit der Lernleistung durch die programmierte Unterweisung, wie sie insbesondere in amerikanischen Veröffentlichungen öfters geäußert wird, liegt nicht vor.
- b) Ungeachtet der Tatsache, daß wir für beide Unterrichtsmethoden eine Korrelation erwartet haben, welche signifikant von Null abweicht, drängt sich bei einer zusammenfassenden Wertung aller Aspekte der vergleichenden theoretischen Betrachtung die Vermutung auf, daß der geprüfte Zusammenhang im herkömmlichen Unterricht größer sei. Die Differenz der beiden mittleren Korrelationskoeffizienten zeigt zwar einen Unterschied, welcher deutlich in Richtung unserer Hypothese weist, er kann jedoch (im Rahmen unserer Stichprobengröße) nicht signifikant abgesichert werden. Beim Vergleich der Grup-

pen $A_{\text{knowl.}/\text{KU}} - B_{\text{knowl.}/\text{PU}}$ zeigt sich zwar eine signifikante Differenz des Zusammenhangs in der von uns erwarteten Richtung. Diese ist allerdings irreführend, da in sie nicht nur Methodeneffekte sondern systematische Gruppenunterschiede mit eingehen, wie der signifikante Unterschied der Gruppen $A_{\text{knowl.}/\text{KU}}$ und $B_{\text{knowl.}/\text{KU}}$ zeigt. Die Analyse des Unterschieds der Korrelationskoeffizienten innerhalb einer Gruppe zeigt, daß die Unterschiede in beiden Fällen in der erwarteten Richtung liegen, in keinem Fall aber signifikant abgesichert sind. Wir können daher zusammenfassend festhalten, daß lediglich eine Tendenz in Richtung unserer Hypothese besteht, d.h. es deutet sich eine höhere Abhängigkeit der Lernleistung von der Intelligenz im Falle des herkömmlichen Unterrichts an.

- c) Die Analyse der beiden Extremgruppen bestätigt das unter Punkt b) berichtete Ergebnis. Würde nämlich die Behauptung, daß der Zusammenhang zwischen Intelligenz und Lernleistung im herkömmlichen Unterricht signifikant höher sei, zutreffen, dann müßte sich gerade die Leistung von Schülern aus dem unteren Bereich unserer Intelligenzverteilung bei Programmierter Unterweisung deutlicher von ihrer Leistung im herkömmlichen Unterricht ($X_{\text{PU}} - X_{\text{KU}}$) abheben, als dies bei Schülern aus dem oberen Bereich unserer Intelligenzverteilung der Fall wäre. Die gefundenen Abweichungen der tatsächlichen mittleren Ränge von den theoretisch zu erwartenden sind zufällig, d.h. daß bezüglich ihrer Intelligenz deutlich voneinander abweichende Gruppen nicht in erkennbar unterschiedlicher Weise durch die Programmierte Unterweisung begünstigt werden. An die-

ser Stelle muß allerdings einschränkend darauf hingewiesen werden, daß dieser Befund ausschließlich für den Bereich unserer empirisch gefundenen Intelligenzverteilung gilt, die insgesamt vom Bevölkerungsdurchschnitt signifikant nach oben abweicht. Zusammenfassend kann festgehalten werden, daß die Programmierete Unterweisung die im herkömmlichen Unterricht durch Intelligenzunterschiede verursachten Leistungsdivergenzen zu nivellieren vermag, daß dies jedoch nicht in dem Maße der Fall ist, wie aufgrund qualitativer Analysen vermutet werden kann.

9.22 Lernzielkategorie "comprehension" (H 12-14)

- a) Der mittlere Zusammenhang (Gruppe A+B) zwischen Intelligenz und Lernleistung erwies sich bei beiden Unterrichtsmethoden als signifikant.
- b) Die Differenz der mittleren Korrelationskoeffizienten ist nichtsignifikant. Im Gegensatz zu 9.21 läßt sich auch keine Tendenz in Richtung unserer Hypothese erkennen.
- c) Der Extremgruppenvergleich zeigt bei beiden Gruppen nichtsignifikante Abweichungen der empirisch ermittelten mittleren Ränge von den theoretisch zu erwartenden Werten.

Zusammenfassend können wir festhalten, daß die Intelligenz auch bei Programmierter Unterweisung in beiden Lernzielkategorien für den Lernerfolg von zentraler Bedeutung ist. Die von Skinner und seinen Anhängern häufig geäußerte Behauptung, daß gerade weniger intelligente Schüler durch die Programmierete Unterweisung besonders begünstigt wür-

den, muß mit Vorsicht aufgenommen werden. Möglicherweise ist jedoch die im Bereich der Lernzielkategorie "knowledge" festgestellte Tendenz pädagogisch relevant, zumal sie doppelt abgesichert ist. Es sei noch darauf hingewiesen, daß es bei Verwendung von streng im Sinne SKINNERS konzipierten Programmen oder bei Zugrundelegung einer größeren Stichprobe unter Umständen möglich wäre, die von SKINNER erwartete verschiedene Wirksamkeit der Programmierten Unterweisung für verschieden intelligente Gruppen signifikant abzusichern.

9.30 Zusammenhang von Lernleistung und der Anzahl "positiver Verstärkungen" im herkömmlichen Unterricht
(H 15-16)

Hier ist der mittlere Korrelationskoeffizient (Gruppe A+B) in beiden Lernzielkategorien nichtsignifikant, beide Werte liegen deutlich im Zufallsbereich. Die Analyse der Einzelgruppen ergibt zwar für die Lernzielkategorie "knowledge" jeweils signifikante Zusammenhänge, diese liegen jedoch in entgegengesetzter Richtung. Dies ist angesichts der Tatsache, daß beide Gruppen von verschiedenen Lehrern unterrichtet wurden, nicht überraschend. So wäre es z.B. denkbar, daß in der Gruppe B die Lehrer bevorzugt leistungsschwache Schüler am Unterricht beteiligten und verstärkten¹⁾. An dieser Stelle wird einmal mehr deutlich, wie

1) Diese Interpretation bietet sich angesichts des negativen Zusammenhangs zwischen der Anzahl an positiven Verstärkungen und der Lernleistung in den beiden Lernzielkategorien an.

gefährlich es im Rahmen von feldexperimentellen Untersuchungen ist, Ergebnisse zu interpretieren, welche nur einfach abgesichert sind und zudem nur auf einer relativ kleinen Stichprobengröße basieren¹⁾. Die Analyse der Extremgruppen deckt sich weitgehend mit den Befunden, welche anhand der Korrelationskoeffizienten gewonnen wurden. In der Gruppe A ist der anhand der Differenzen $X_{PU} - X_{KU}$ errechnete mittlere Rang in beiden Lernzielkategorien bei den im herkömmlichen Unterricht extrem selten verstärkten Schülern zwar höher (d.h. numerisch geringer), die Werte liegen jedoch deutlich im zufälligen Bereich, d.h. es wurde nicht annähernd das 5%-Niveau erreicht, so daß bei Beibehaltung der Nullhypothese ein Fehler vom Typ β äußerst unwahrscheinlich ist. Bei der Gruppe B ist der mittlere Rang der im herkömmlichen Unterricht extrem häufig verstärkten Schüler sogar höher. Würde der Häufigkeit an positiven Verstärkungen im Rahmen des herkömmlichen Unterrichts bezüglich der Lernleistung diejenige Bedeutung zukommen, welche ihr weitgehend zuerkannt wird, müßte sich dies bei den im herkömmlichen Unterricht extrem selten verstärkten Schülern dahingehend auswirken, daß gerade sie durch die programmierte Unterweisung besonders begünstigt würden. Wir müssen allerdings an dieser Stelle darauf

1) Dies gilt insbesondere für Korrelationskoeffizienten, welche bei Verwendung einer relativ kleinen Stichprobe errechnet werden. Gerade in dieser Hinsicht wimmelt es im Forschungsbereich der programmierten Unterweisung an Fehlinterpretationen. Ein diesbezüglich besonders krasser Fall findet sich bei CORRELL, welcher davon ausgeht, daß bei programmierter Unterweisung "weniger Schulangst mit im Spiele" ist als im Falle des herkömmlichen Unterrichts. Im Anschluß an diese Feststellung schreibt der Autor: "Tatsächlich ließ sich ... feststellen, daß dies der Fall war: Die Interkorrelation war bei der Kontrollgruppe 0.21 und bei der Experimentiergruppe 0.19" (W. CORRELL: "Programmiertes Lernen und schöpferisches Denken", München 1965, S. 51).

hinweisen, daß die Beibehaltung der Nullhypothese kein Beweis dafür sein kann, daß die untersuchten und nicht aufgefundenen Zusammenhänge *n i c h t* vorhanden sind, streng genommen kann nur festgehalten werden, daß solche Zusammenhänge nicht aufgewiesen werden konnten. Angesichts der *e x t r e m* geringen Zusammenhänge muß die gegenteilige Behauptung jedoch mit äußerster Vorsicht aufgenommen werden, zumal darüber hinaus ähnlich gelagerten Untersuchungen in den Vereinigten Staaten nicht der Nachweis gelang, daß Programme mit eingebauter sofortiger Verstärkung eine signifikant höhere Lernleistung bewirken als Programme desselben Inhalts und didaktischen Aufbaus, welche auf die Verstärkungen verzichteten¹⁾. Zusammenfassend meinen wir daher festhalten zu können, daß die zentrale Bedeutung, welche der Häufigkeit an Verstärkungen im Rahmen der Diskussion der Programmierten Unterweisung zuerkannt wird, kritisch überdacht werden sollte. Die ungenügende Lehrwirksamkeit eines herkömmlichen Unterrichts kann nicht mit der relativ geringen Zahl der pro Schüler entfallenden Verstärkungen begründet werden²⁾. Umgekehrt kann die Überlegenheit der Programmierten Unterweisung nicht mit der relativ großen Anzahl an Verstärkungen begründet werden. Vorsichtiger formuliert: Die gegenteilige, weit verbreitete Ansicht wurde für schulisches Lernen der Art, wie wir es untersucht haben, nie empirisch verifiziert. Hinweise auf die SKINNERsche Lerntheorie können nicht als Verifikation gelten, da die direkte Übertragung dieser Lerntheorie auf die komplexeren

1) Vgl. 2.24

2) Vgl. B.F. SKINNER: "Die Wissenschaft vom Lernen und die Kunst des Lehrens", in: W. CORRELL (Hrsg.): "Programmiertes Lernen und Lehrmaschinen", Braunschweig 1966, S. 74 f.

Bedingungen des schulischen Lernens zumindest problematisch ist¹⁾. Wir wollen damit in keiner Weise die Aussagekraft dieser Lerntheorie sowie die Bedeutung der Verstärkung für menschliches Lernen leugnen, uns liegt lediglich an dem Hinweis, daß bei schulischen Lernprozessen der Art, wie wir sie untersucht haben, der Häufigkeit an Verstärkungen längst nicht der Stellenwert zukommen kann, der ihr heute noch vielerorts zuerkannt wird²⁾.

-
- 1) In anderem Zusammenhang äußert TOMAN über die Schwäche der lerntheoretischen Ansätze: "Lerntheoretiker sind oft groß in der Anwendung ad hoc ihrer Modelle auf die komplexere Wirklichkeit. Dabei ist das, was ihnen an dieser Wirklichkeit einleuchtet, ein beschränkter und manchmal geradezu kümmerlicher Ausschnitt derselben. Eine der smartesten Anwendungen hat SKINNER in seiner Schilderung einer utopischen Gesellschaft geliefert (SKINNER 1948), in der positive Verstärkung, operative Konditionierung ... und Stundenpläne der Verstärkungen ... die Hauptprinzipien darstellen, nach denen die Menschen von "Walden II" in ihre Gesellschaft hineinwachsen" (W. TOMAN: "Motivation, Persönlichkeit, Umwelt", Göttingen 1968, S. 60 f.).
 - 2) Abgesehen von CROWDER (N.A. CROWDER: "Über die Unterschiede zwischen linearem und verzweigtem Programmieren", in: W. CORRELL (Hrsg.): "Zur Theorie und Praxis des programmierten Lernens", Darmstadt 1969) haben insbesondere LUMSDAINE im amerikanischen (A.A. LUMSDAINE: "Einige theoretische und praktische Probleme beim programmierten Unterricht", in: W. CORRELL (Hrsg.): "Zur Theorie und Praxis des programmierten Lernens", Darmstadt 1969) und L.N. LANDA u.a. (L.N. LANDA, A.M. ORLOVA, G.G. GRANIK: "Prinzipien des programmierten Unterrichts der russischen Sprache", in: W. CORRELL (Hrsg.): "Zur Theorie und Praxis des programmierten Lernens", Darmstadt 1969) im russischen Bereich auf die Überbewertung des Prinzips der Verstärkung im Kontext der Diskussion über die programmierte Unterweisung hingewiesen.

9.40 Zusammenhang von Lernleistung und "unterrichtsbezogener Schüleraktivität" (H 17-18)

Die beiden mittleren Korrelationskoeffizienten sind nichtsignifikant. Die Differenzen $T_1 - T_{\text{pop}}$ der Extremgruppen entsprechen weitgehend den unter 9.30 berichteten Werten, d.h. bei der Gruppe A hat die im herkömmlichen Unterricht extrem selten unterrichtsbezogene Aktivität aufweisende Gruppe bei beiden Lernzielkategorien zwar jeweils einen höheren mittleren Rang, diese liegen jedoch noch sehr deutlich im zufälligen Bereich. Bei der Gruppe B ergab sich sogar für diejenige Gruppe, welche im herkömmlichen Unterricht eine extrem hohe unterrichtsbezogene Aktivität aufweist, ein signifikant höherer mittlerer Rang. Bezüglich der widersprüchlichen Richtungen unserer Ergebnisse gelten die unter 9.30 gemachten Ausführungen. Wenn der aktiven Formulierung einer Antwort diejenige Bedeutung zukommen würde, welche ihr im Rahmen der Diskussion über die Programmierte Unterweisung häufig zuerkannt wird, dann müßte erwartet werden, daß Schüler, welche im herkömmlichen Unterricht extrem selten ausdrücklich geäußerte unterrichtsbezogene Aktivität aufweisen, durch die Programmierte Unterweisung besonders begünstigt würden. Es scheint angesichts unserer Ergebnisse jedoch äußerst fraglich zu sein, ob diese Annahme tatsächlich zutrifft. Die Befunde unserer Arbeit und von ähnlich gelagerten Untersuchungen in den Vereinigten Staaten weisen nicht in diese Richtung¹⁾. In unserem Fall können wir mit Sicherheit davon ausgehen, daß die festgestellte generelle höhere Lehreffektivität der Programmierten Unterweisung nicht auf die Konstruktions-

1) Vgl. 2.23

antwortmethode zurückgeführt werden kann. Es ist daher auch nicht empfehlenswert, im Rahmen der Unterrichtsdifferenzierung Programmgruppen zu bilden, welche sich durch gering ausgeprägtes Antwortverhalten auszeichnen, da sie vermutlich durch die Programmierte Unterweisung nicht mehr begünstigt würden als die übrigen Klassenmitglieder. Weiter muß das im Rahmen der Programmkonstruktion häufig zu beobachtende Beharren auf der Konstruktionsantwortmethode neu überdacht werden.

9.50 Der Zusammenhang von "konzentrativer Arbeitshaltung und der Lernleistung in den beiden Unterrichtsmethoden (H 19-20)

Sowohl im herkömmlichen als auch im programmierten Unterricht ergab sich auf der Basis der partiellen mittleren Korrelationskoeffizienten (Gruppe A+B) ein signifikanter Zusammenhang zwischen der Lernleistung und der Fähigkeit zu konzentriertem, eigenständigem Arbeiten. Die Differenz der mittleren partiellen Korrelationskoeffizienten ist bei der Lernzielkategorie "knowledge" beinahe signifikant. Bei der Lernzielkategorie "comprehension" liegt die Differenz deutlich im zufälligen Bereich. Dieser Befund wird durch die Analyse der Einzelgruppen bestätigt. Bei beiden Gruppen zeigt sich, daß der Zusammenhang zwischen der Lernleistung und der Fähigkeit zu konzentriertem, eigenständigem Arbeiten bei der Lernzielkategorie "knowledge" im herkömmlichen Unterricht deutlich größer ist als bei Programmierter Unterweisung, in jedem Fall allerdings nicht auf dem 5%-Niveau abgesichert werden kann. Obwohl der Unterschied der Ausprägung des Zusammenhangs in allen Fällen noch zufällig ist, kann trotzdem eine deutlich er-

kennbare Tendenz in Richtung unserer Hypothese für die Lernzielkategorie "knowledge" konstatiert werden¹⁾. Dies kann als eine, wenn auch nur unvollkommen abgesicherte, Bestätigung der Auffassung von SKINNER interpretiert werden, derzufolge Schüler, die im herkömmlichen Unterricht als unkonzentriert arbeitend auffallen, durch die Programmierte Unterweisung besonders begünstigt werden²⁾. Nach SKINNER handelt es sich dabei insbesondere um Schüler, deren Arbeitshaltung und Lernleistung durch das Eingebettetsein der Lernprozesse in den sozialen Kontext negativ beeinträchtigt werden³⁾. Wenn mancherorts von Schulpraktikern angenommen wird, die Programmierte Unterweisung eigne sich für "unkonzentriert" arbeitende Schüler nicht, da diese nicht dazu in der Lage seien, ohne Lehrerhilfe zu arbeiten, so wird dabei außer acht gelassen, daß bei Vorliegen einer unkonzentrierten Arbeitshaltung im herkömmlichen Unterricht nicht automatisch geschlossen werden kann, daß dies für andere Arbeitsbedingungen, wie es z.B. die Programmierte Unterweisung darstellt, auch gilt. Ein Blick auf unsere Form des herkömmlichen Vergleichsunterrichts macht weiter deutlich, daß die Schüler hier viel eher abgelenkt und unterbrochen werden, und daher auch einer ausgeprägteren Arbeitshaltung bedürfen um sich auf diejenigen Komponenten des vielfältigen Reizgeschehens im Klassenzimmer zu konzentrieren, welche der Stoffaneignung förderlich sind. Im herkömmlichen Unterricht befindet sich jeder Schüler in einem permanenten Kontakt mit Lehrer und Schülern, wobei insbeson-

-
- 1) Wir haben einigen Grund zu der Vermutung, daß bei Zugrundelegung einer größeren Stichprobe unsere Hypothese signifikant abgesichert werden könnte.
 - 2) Wir müssen an dieser Stelle ausdrücklich darauf hinweisen, daß der häufig geäußerte und z.T. empirisch abgesicherte Hinweis, die Programmierte Unterweisung sei für die Schüler anstrengender als der herkömmliche Unterricht, unserem Tendenzbefund nicht widerspricht.
 - 3) Vgl. B.F. SKINNER: "Warum Lehrer versagen", in: W. CORRELL (Hrsg.): "Zur Theorie und Praxis des programmierten Lernens", Darmstadt 1969, S. 93 f.

dere - auch bei "gut geführten Klassen" - seitens der Mitschüler mit sehr viel ablenkenden Reizen gerechnet werden muß. Demgegenüber ist der Aufwand, welcher zur Abschirmung störender Umweltreize notwendig ist, bei Programmierter Unterweisung weitaus geringer, da störende Kontakte hier aufgrund der äußeren Bedingungen auf ein Minimum reduziert werden. Schließlich soll auch noch daran erinnert werden, daß diejenigen Schüler, welche vom Lehrer als "unkonzentriert" arbeitend bezeichnet werden, häufig mit denen identisch sind, welche sich im Lehrerunterricht nicht zurechtfinden, sei es, daß die Stoffdarbietung zu schnell oder zu langsam¹⁾ erfolgt, sei es, daß ein Schüler während einer bestimmten Unterrichtsphase in einer ganz bestimmten Richtung "fixiert" ist und durch Beiträge und Fragen seitens der Lehrer und Mitschüler, welche in eine andere Richtung zielen, abgelenkt bzw. verwirrt wird. Die von uns als beinahe signifikant aufgewiesene Differenz der mittleren Korrelationskoeffizienten kann als Hinweis dafür dienen, daß die Arbeitshaltung ein Kriterium ist, welches im Rahmen des differenzierten Einsatzes der Programmierten Unterweisung zur Gruppeneinteilung mit herangezogen werden kann.

9.60 Der Zusammenhang zwischen der Einstellung der Schüler zu Lehrer und Schule und der Lernleistung (H 21-22)

Bezüglich der Schuleinstellung ergaben sich folgende Werte: Die mittleren Korrelationskoeffizienten (Gruppe A+B)

1) An dieser Stelle sei an die vielfältigen Verhaltensstörungen erinnert, welche im Zusammenhang mit der Über- bzw. Unterforderung entstehen können. Es gehört zum gesicherten Bestand der einschlägigen Unterrichtsbeobachtung, daß über- bzw. unterforderte Schüler durch unkonzentriertes Verhalten auffallen.

und deren Differenzen sind alle nichtsignifikant. Die Analyse des Zusammenhangs zwischen der Lernleistung und der Einstellung zum Lehrer zeigt, daß auch hier sämtliche mittleren Korrelationskoeffizienten nichtsignifikant sind. Dies gilt auch für die Differenzen derselben. Bei den Einzelgruppen zeigt sich weitgehend dasselbe Bild. Demgegenüber wurde bislang häufig geäußert und auch von uns vermutet, daß Schüler mit negativer Schul- und Lehrereinstellung mittels der Programmierten Unterweisung aufgrund der sozialen Entlastung der Lernsituation besonders effizient zu unterrichten seien. Diese Annahme basiert allerdings ausschließlich auf einer qualitativen Analyse der Wirkbedingungen der Programmierten Unterweisung. Angesichts des Tatbestands, daß sich in keiner der beiden Lernzielkategorien, weder bei Zugrundelegung eines bereits erprobten und auf dem Markt befindlichen Verfahrens noch bei Verwendung von numerischen Schätzskalen, in keiner unserer beiden Gruppen auch nur annähernd signifikante Unterschiede nachweisen ließen¹⁾, meinen wir feststellen zu können, daß die oben geschilderten Annahmen einer Überprüfung bedürfen. Da unsere Schüler in ihrer Gesamtheit keinesfalls eine ausgeprägt negative Einstellung Schule und Lehrer gegenüber haben (dies zeigt besonders deutlich die bei beiden Gruppen beobachtete signifikante Bevorzugung des Lehrerunterrichts), können wir genau genommen nur für einen mittleren Einstellungsbereich eine Aussage machen. So wäre es durchaus denkbar, daß sich im Falle des Rückgriffs auf Schüler mit extrem negativer Einstellung die von uns hypothetisch angenommenen Differenzen

1) Diese Aussage bezieht sich ausschließlich auf die mittleren Koeffizienten.

ergeben hätten. Aber selbst wenn diese Annahme zutreffen würde, wäre sie für die Schulpraxis von relativ geringer Bedeutung, da angesichts des immer mehr praktizierten sozialintegrativen Unterrichtsstils extrem negative Einstellungen immer seltener werden und somit die Voraussetzungen für eine Differenzierung innerhalb des Klassenunterrichts im Hinblick auf Einstellungen entfallen. Es könnte weiter vermutet werden, daß es zwar im Falle des Vorliegens von nur einem Merkmal (z.B. schlechte Lehrereinstellung) nicht möglich ist, die vermuteten Effekte nachzuweisen, daß diese sich jedoch bei gleichzeitigem Vorliegen mehrerer Merkmale mit entsprechender Ausprägung nachweisen lassen. Wir sind dieser Frage nachgegangen, konnten jedoch auch bei Zugrundelegung von Extremgruppen, welche durch mehrere Merkmale gekennzeichnet sind, keine Ergebnisse erhalten, welche von den oben berichteten Befunden in charakteristischer Weise abweichen¹⁾. Die erhaltenen Werte liegen auch unter diesen Bedingungen für beide Gruppen und beide Lernzielkategorien deutlich im zufälligen Bereich. Die Beträge $|T_1 - T_{\text{pop}}|$ sind dabei so gering, daß trotz der durchgehend in Richtung der Erwartung liegenden mittleren Ränge nicht einmal von einer Tendenz gesprochen werden kann. Angesichts der durchgängigen Übereinstimmung unserer Befunde muß davon ausgegangen werden, daß der Einfluß der Einstellung auf die Lernleistung bei beiden Unterrichtsmethoden vermutlich längst nicht die Wirkung hat, wie das vielerorts angenommen wird, d.h. daß entsprechende Gruppendifferenzierungen, wie sie in der Schulpraxis schon vorgenommen werden, wahrscheinlich nicht sinnvoll sind. Wir müssen dabei ausdrücklich darauf hinweisen, daß unsere Aussagen zunächst ausschließlich für Lernleistungen gelten, welche auf Lernprozessen basieren, die sich über einen relativ kurzen Zeitraum erstrecken.

- 1) Zur Klärung dieser Frage haben wir Extremgruppen gebildet, welche sich dadurch auszeichneten, daß sie wenigstens bezüglich zweier Kriterien dem oberen bzw. unteren Fünftel unserer empirisch gefundenen Verteilungen angehörten. Zusätzlich zu den Kriterien "Einstellung zur Schule" und "Einstellung zum Lehrer" verwendeten wir noch das Ausmaß an Schulangst. Dies ermittelten wir mit Hilfe einer modifizierten Form des AAT ("Anxiety in academic achievement situations" von R. ALPERT und R.N. HABER), welche von KEIL und SADER zusammengestellt wurde (W. KEIL und M. SADER: "Situationsspezifische Einflüsse auf die Leistungsmotivationsmessung nach Heckhausen", in: Ztschr. f. exp. u. ang. Psych., 1, 1968, S. 103). Bei den solchermaßen gebildeten Extremgruppen ergaben sich folgende absoluten Differenzen $|T_1 - T_{pop}|$: $A_{knowl.} : 8$; $A_{compr.} : 6,5$; $B_{knowl.} : 3,5$; $B_{compr.} : 13$. In diesem Zusammenhang sei der Vollständigkeit halber noch erwähnt, daß das Ausmaß an Schulangst mit der Lernleistung in beiden Lernzielkategorien zwar signifikant (negativ) korreliert, daß aber die Differenzen zwischen den beiden Unterrichtsmethoden zufällig sind (die Berechnung der mittleren Korrelationskoeffizienten ergab folgende Werte: $r_{M(A+B)/knowl./KU} = -0,24$; $r_{M(A+B)/knowl./PU} = -0,23$; $r_{M(A+B)/compr./PU} = -0,19$; $r_{M(A+B)/compr./KU} = -0,14$ ($r_{krit\ 5\%} = 0,183$; $r_{krit\ 1\%} = 0,234$)).

Weiter können wir festhalten, daß die von uns festgestellte generelle höhere Lehreffektivität der Programmierten Unterweisung nicht mit dem Prinzip der sozialen Entlastung der Lernsituation erklärt werden kann¹⁾.

9.70 Zusammenfassung

- a) Abschließend stellt sich die Frage, wodurch die von uns festgestellte durchgängig höhere Lehreffektivität der Programmierten Unterweisung verursacht wurde. Wir haben gezeigt, daß die Häufigkeit des Vorkommens verstärkender Reize und die Häufigkeit "unterrichtsbezogener Schüleraktivität" im herkömmlichen Unterricht bezüglich dessen "Lehreffektivität" nicht diejenige Bedeutung haben, welche ihnen im Anschluß an SKINNER zuerkannt wird. Daraus kann der Schluß abgeleitet werden, daß im Rahmen unserer Untersuchung die Überlegenheit der Programmierten Unterweisung nicht mit häufigerer Verstärkung und der Konstruktionsantwortmethode erklärt werden kann. Weiter kann das Prinzip der sozialen Entlastung der Lernsituation in unserem Fall nicht herangezogen werden, da auch hier nur ein deutlich zufälliger Zusammenhang mit der Lernleistung in beiden Unterrichtsmethoden festgestellt werden konnte. Auch die in Form von Tendenzen festgestellte Begünstigung der weniger intelligenten Schüler kann nicht

1) Dafür spricht auch der Tatbestand, daß die Differenz des Zusammenhangs zwischen Schulangst und der Lernleistung bei den beiden Unterrichtsmethoden zufällig ist (siehe oben).

als ausreichendes Erklärungsprinzip herangezogen werden, dasselbe gilt für das Kriterium "konzentrierte Arbeitshaltung". Wir können also mit einiger Wahrscheinlichkeit die Vermutung äußern, daß die Ursachen für die festgestellte Überlegenheit der Programmierten Unterweisung nicht primär in der Begünstigung definierter Teilgruppen unserer Gesamtstichprobe liegen, sondern vermutlich in erster Linie in Funktionsprinzipien, welche für alle bzw. einen Großteil unserer Schüler eine höhere Lehr-effektivität bewirkten. Einmal scheint uns hier das Prinzip des individuellen Lerntempos von zentraler Bedeutung zu sein. Insbesondere langsam arbeitende Schüler, aber auch Schüler mit fluktuierender Aufmerksamkeitshaltung verpassen im herkömmlichen Unterricht laufend den Anschluß und sind, da die unterrichteten Inhalte meistens aufeinander aufbauen, schon nach relativ kurzer Zeit nicht mehr in der Lage, allen Ausführungen des Lehrers zu folgen und richtig (aktiv oder innerlich) zu antworten, da sie nur einen Teil der zur jeweiligen Aufgabenlösung notwendigen Verhaltensweisen gelernt haben. An dieser Stelle könnte nun vermerkt werden, daß in diesem Fall auch die Häufigkeit der direkten Verstärkungen und die Wahrscheinlichkeit der kollektiven Verstärkungen verringert würden und daß somit die Verstärkung doch entscheidend für den Erfolg der Lehrbemühungen sei. Das heißt doch wohl aber, daß nicht primär die H ä u f i g k e i t der Verstärkungen sondern deren systematische und kontinuierliche Verabreichung im Rahmen des gesamten Lehr-

konzepts von Bedeutung ist¹⁾. Damit sind wir beim zweiten entscheidenden Vorteil der Programmierten Unterweisung: sie gewährt weit mehr als der herkömmliche Unterricht eine systematische und kontinuierliche Verhaltensformung. Der Schüler wird kontinuierlich von Lernziel zu Lernziel geführt. Dabei ist nicht entscheidend, daß er möglichst oft, sondern daß er an den entscheidenden Stellen verstärkt wird. Wir haben im Rahmen unserer Unterrichtsbeobachtung häufig registriert, daß einzelne Schüler sehr oft verstärkt wurden, daß diese Verstärkungen jedoch ausschließlich im Verlauf der Lösung einzelner Aufgaben verabreicht wurden²⁾. D.h. daß diese Schüler punktuell optimal im Sinne SKINNERS unterrichtet wurden, weit besser, als dies das verwendete Programm vermocht hätte, daß aber eine systematische Verhaltensformung über a l l e untergeordneten Lernziele hinweg bis zum endgültig angestrebten Verhalten in keinem Fall auch nur annähernd vorlag. Wir meinen daher einen besonderen Vorzug der Programmierten Unterweisung darin zu sehen, daß diese i.allg. eine viel sorgfältigere und zielgerichtete Lernstrategie aufweist. Mit anderen Worten: Unterrichtsprogramme sind didaktisch besser durchdacht und

-
- 1) An dieser Stelle sei noch einmal ausdrücklich darauf verwiesen, daß SKINNER bei der Kritik des herkömmlichen Unterrichts u.a. ausgedrückt auf das Problem der Häufigkeit der verabreichten Verstärkungen hingewiesen hat.
 - 2) So konnten wir z.B. beobachten, daß ein Schüler während der Lösung einer einfachen Bruchpotenz 6 verstärkende Reize seitens des Lehrers erhielt. Diese bezogen sich auf kleinste Teilschritte der Lösung. Eine entsprechende Lösung wird im Kontext des Lehrprogramms lediglich e i n m a l verstärkt.

schlüssiger aufgebaut, wobei die ausschließliche Konfrontation eines Schülers mit einem Programm sich dahingehend positiv auswirkt, daß der Schüler die von der Lernstrategie geplante Abfolge der Lernschritte vollständig durchschreitet. Der herkömmlich unterrichtende Lehrer, welcher im Hinblick auf die höhere Lehr-effektivität der Programmierten Unterweisung seinen eigenen Unterricht verbessern will, wird sich insbesondere um eine Verbesserung seiner Lehrstrategie zu bemühen haben. Es ist zwar ein wenig übertrieben, wenn hier festgehalten wird, daß die detaillierten Stationen des Wegs zum endgültigen Lernziel im herkömmlichen Unterricht i.allg. weitgehend unbekannt sind und man schließlich beinahe zufällig dorthin gelangt, für die Mikrostruktur des herkömmlichen Unterrichts dürfte dies aber häufig weitgehend zutreffen.¹⁾ Die größere Beweglichkeit des herkömmlichen Unterrichts erweist sich hier als Nachteil, da der Vorteil derselben i.allg. ausschließlich denjenigen Schülern zugute kommt, welche in einer bestimmten unterrichtlichen Situation direkten Kontakt mit dem Lehrer aufnehmen, die übrigen Schüler können durch diese Beweglichkeit sogar häufig irritiert werden. Wir können also zusammenfassend die Vermutung äußern, daß aufgrund unserer Ergebnisse in erster Linie das Prinzip des individuel-

1) In diesem Zusammenhang muß daran erinnert werden, daß sich unser herkömmlicher Unterricht aufgrund der teilweise erfolgten Festlegung der Mikrostruktur vorteilhaft vom üblichen Lehrerunterricht abhebt, d.h. daß unter nichtexperimentellen Bedingungen der Programmierten Unterweisung noch eher Vorteile eingeräumt werden müssen.

len Arbeitstempos und die sorgfältiger und schlüssiger konzipierten und regelmäßig verabreichten Lernschritte als Ursache für die höhere Lehreffektivität der Programmierten Unterweisung in Frage kommen. Dies gilt für beide der von uns untersuchten Lernzielkategorien.

- b) Die unter Einbezug von programmiertem Arbeitsmaterial vorgenommenen Gruppendifferenzierungen sollten nicht im Hinblick auf die von uns untersuchten adressatenbezogenen Maßkriterien "Einstellung zu Lehrer und Schule" erfolgen, da nicht erwartet werden kann, daß in diesen Fällen die programmiert unterrichtete Gruppe mit besserem Erfolg abschneiden würde, als dies im herkömmlichen Unterricht der Fall wäre. Für Schüler mit geringer Intelligenz und wenig ausgeprägter "konzentrativer Arbeitshaltung" k a n n im Bereich der Lernzielkategorie "knowledge" eine solche Gruppenaufteilung in Betracht gezogen werden. Ungeachtet dieser Möglichkeit sind wir jedoch der Ansicht, daß der Gruppendifferenzierung weniger adressatenbezogene Maßkriterien sondern eher didaktische Erwägungen zugrundegelegt werden sollten¹⁾.
- c) Das weit verbreitete Vorurteil, daß der herkömmliche Unterricht besser als die Programmierte Unterweisung geeignet sei, die Voraussetzungen für die Lösung von Aufgaben zu schaffen, welche nicht direkt behandelt worden sind und Transferleistung verlangen, konnte widerlegt werden. Die höhere Lehreffektivität der Programmierten Unterweisung wirkt sich auch in diesem Bereich aus.

1) Vgl. G. SCHRÖTER: "Objektivierung des Unterrichts", Braunschweig

Literatur

- AEBLI, H.: "Programmierte Ausbildung in didaktischer Sicht", in: Deutsche Schule, Hannover 58, 1966
- AEBLI, H.: "Grundformen des Lehrens", München 1968
- AEBLI, H.: "Psychologische Didaktik", Stuttgart 1968
- AMTHAUER, R.: "Intelligenz-Struktur-Test", Göttingen 1955
- ANGELL, D. and LUMSDAINE, A.A.: "The Effects of Prompting Trials and Partial-Correction Procedures on Learning by Anticipation", San Mateo, California: American Institute for Research, 1961
- ANGELL, G.W.: "The effect of immediate knowledge of quiz results on final examination scores in freshmen chemistry", in: Journ. educ. Res., Jan. 1949, 42
- ANNET, J.: "The Role of Knowledge of Results in Learning: A Survey", in: J.P. De CECCO (Hrsg.): Educational Technology", New York 1964
- AUBLE, D. and MECH, E.: "Response Strength in a Classroom Task Related to Forward Delay in Reinforcement", in: Journ. educ. Psychol., 45, 1954
- BARLOW, J.A.: "Conversational Chaining in Teaching Machine Programs", Earlham College, Richmond 1960

- BLANKERTZ, H.: "Theorien und Modelle der Didaktik",
München 1969
- BLOOM, B.S. (Hrsg.): "Taxonomy of educational objectives", New York 1956
- BLYTH, J.W.: "Teaching Machines and human beings",
in: A.A. LUMSDAINE and R. GLASER (Ed.):
"Teaching Machines and Programmed Learning",
Washington 1960
- BRAUNE, H.: "Untersuchung über die Wirkungsweise verschiedener Bestätigungen (reinforcement; confirmation) in einem verzweigten Programm unter Berücksichtigung differentiell-psychologischer Lernvariablen", in: ZeF (Zeitschrift für erziehungswissenschaftliche Forschung),
3, 1969, 3 und 4
- BRUNER, J.S.: "The art of discovery", in: Harv. Ed. Rev.,
1961, 31
- BYRAN, G.L. and RIGNEY, J.W.: "An evaluation of a method of shipboard training in operations knowledge",
in: J.P. De GECCO (Ed.): "Educational Technology", New York 1964
- CALL, W.A. Mc.: "How to Experiment in Education", New
York 1923
- CAMPBELL, D.T. and STANLEY, J.C.: "Experimental and Quasi-Experimental Designs for Research on Teaching", in: N.L. GAGE (Ed.): "Handbook of Research on Teaching", Chicago 1965

- CASSEL, R. and TELLOM, W.L.: "A preliminary evaluation of programmes instruction with students of high ability", in: Psych. Rep., 1962, 10 (1).
- CHOMSKY, N.: "Review of B.F. Skinner, Verbal Behavior", Language 35, 1959
- CORRELL, W.: "Programmiertes Lernen und schöpferisches Denken", München 1965
- CORRELL, W.: "Das pädagogisch-psychologische Problem der Beziehung zwischen Lehrstil und Lernleistung", in: Th. HERMANN (Hrsg.): "Psychologie der Erziehungsstile", Göttingen 1966
- CORRELL, W.: "Die Information im Unterrichtsprogramm", in: B.F. SKINNER und W. CORRELL: "Denken und Lernen", Braunschweig 1967
- CORRELL, W.: "Pädagogische Verhaltenspsychologie", München-Basel 1967
- CORRELL, W.: "Lernstörungen beim Schulkind", Donauwörth 1968
- CORRELL, W.: "Zur Theorie und Praxis des programmierten Lernens", Darmstadt 1969
- COULSON, J.E. and SILBERMAN, H.F.: "Effects of Three Variables in a Teaching Machine", in: Journ. of Educ. Psychol., 51, 1960
- COULSON, J.E. and SILBERMAN, H.F.: "Results of initial experiment in automated teaching", in: A.A. LUMSDAINE and R. GLASER (Ed.): "Teaching machines and programmed learning", Washington 1960

- CROWDER, N.A.: "Unterschiede zwischen linearem und verzweigtem Programmieren", in: W. CORRELL (Hrsg.): "Zur Theorie und Praxis des programmierten Lernens", Darmstadt 1969
- CSANYI, A.P.; GLASER, R.; REYNOLDS, H.: "Programmiermethode und Antwortart bei einem Experiment zur Schulung der Aussprache mit visuellen Hilfen", in: W. CORRELL (Hrsg.): "Zur Theorie und Praxis des programmierten Lernens", Darmstadt 1969
- CUBE, F.v.: "Kybernetische Grundlagen des Lernens und Lehrens", Stuttgart 1968
- DIETRICH, G.: "Bildungswirkungen des Gruppenunterrichts", München 1969
- DIETRICH, G.: "Die Auswirkungen von Unterrichtsstilen auf das Leistungs- und Sozialverhalten von Schülern", in: Blätter für Lehrerfortbildung. "Das Seminar", 2, 1971
- DRALLE, W.: "Erfahrungen mit einem Jahrgangsprogramm", in: B. ROLLET (Hrsg.): "Praxis und Theorie des programmierten Unterrichts", Stuttgart-München, 1970
- ECKEL, K.: "Die Bedeutung des Klasseneffekts für die Schulpädagogische Forschung", in: pl, 3, 1969
- ECKEL, K.: "Von der empirischen zur erfahrungswissenschaftlichen Pädagogik", in: pl, 4, 1970

- EIGEN, L.D.: "Probleme der Forschung im Bereich des Programmierten Unterrichts", in: L.J. ISSING (Hrsg.): "Die Programmierte Unterweisung in den USA heute", Weinheim 1967
- EIGLER, G.: "Prüfung von Lehrprogrammen", in: ZeF (Zeitschrift für erziehungswissenschaftliche Forschung), 2, 1968, 3
- EVANS, J.L.: "Programing in Mathematics and Logic", in: R. GLASER (Ed.): "Teaching Machines and Programed Learning II", New York 1965
- EVANS, J.L.; GLASER, R. and HOMME, L.E.: "A Preliminary Investigation in the Properties of Verbal Learning Sequences of the 'Teaching Machines' Type", in: R. Glaser und A.A. LUMSDAINE (Ed.): "Programmed Learning and Teaching Machines", New York 1960
- EVANS, J.L.; GLASER, R. and HOMME, L.E.: "The Ruleg System for the Construction of Programmed Verbal Learning Sequences", Department of Psychology, Univ. of Pittsburgh, 1960
- EYFERTH, K.: "Lernen als Anpassung des Organismus durch bedingte Reaktionen", in: R. BERGIUS (Hrsg.): "Lernen und Denken", Hb. d. Psych., Bd. 1,2, Göttingen 1964
- EYFERTH, K.: "Methoden zur Erfassung von Erziehungsstilen", in: Th. HERMANN (Hrsg.): "Psychologie der Erziehungsstile", Göttingen 1966

- FERSTER, C.B. and SKINNER, B.F.: "Schedules of reinforcement", New York 1957
- FIEBIG, H. und JUNGHANS, H.: "Rechnen mit Potenzen", Braunschweig o.J.
- FIEBIG, H. und JUNGHANS, H.: "Rechnen mit Wurzeln", Braunschweig o.J.
- FIEBIG, H. und JUNGHANS, H.: "Rechnen mit Logarithmen", Braunschweig o.J.
- FILBY, Y.: "Teaching Machines. A Review of Theory and Research", Copenhagen 1961
- FIPFINGER, F.: "Intelligenz und Schulleistung bei 9- und 10jährigen Volksschülern (Untersuchungsergebnisse)", in: K.-H. INGENKAMP (Hrsg.): "Schulkonflikt und Schülerhilfe", Weinheim 1965
- FOLTZ, C.I.: "Lehrmaschinen", Weinheim 1965
- FOPPA, K.: "Lernen, Gedächtnis, Verhalten", Köln 1966
- FRANK, H.: "Kybernetische Grundlagen der Pädagogik", Baden-Baden 1962
- FRANK, H. (Hrsg.): "Lehrmaschinen in kybernetischer und pädagogischer Sicht", 4 Bde., Stuttgart-München, 1963, 1964, 1965, 1966
- FREUDENSTEIN, R.: "Der Programmierte Unterricht und seine lerntheoretischen Grundlagen", in: W. KLAFKI (Hrsg.): "Erziehungswissenschaft 2", Frankfurt 1970
- FRY, E.B.: "A study of teaching machine response modes", in: A.A. LUMSDAINE and R. GRASER (Ed.): "Teaching Machines and Programmed Learning", Washington 1960

- FRY, E.B.: "Teaching Machines and programed instruction. An Introduction", New York 1963
- GAGNÉ, M.: "The analysis of instructional objectives for the design of instruction", in: R. GLASER (Ed.): "Teaching machines and programmed learning II", New York 1965
- GAGNÉ, R.M. and DICK, W.: "Learning Measure in a Self-Instructional Program", in: J.P. De CECCO (Ed.): "Educational Technology", New York 1964
- GAVURIN, E.L. and DONAHUE, V.A.: "Logical Sequence and Random Sequence in Teaching Machine Programs", in: "Automated Teaching Bulletin, I, 1961
- GEISLER, E.E.: "Herbarts Theorie des 'kleinsten Schrittes' und der Programmierte Unterricht", in: Westermanns Pädagogische Beiträge, 18, 1966
- GILBERT, T.F.: "Mathetics. The Technology of Education, I. The General System", TOR Education, Inc., New York 1961
- GLASER, R.: "Some research problems in automated instruction: instructional programming and subject-matter structure", in: E. COULSON (Ed.): "Programmed Learning", New York 1962
- GLASER, R.: "Lernen und die Technologie des Unterrichts", in: W. CORRELL (Hrsg.): "Zur Theorie und Praxis des programmierten Lernens", Darmstadt 1969

- GLASER, R.; REYNOLDS, J.H.; FULLICK, M.G.: "Programmierter Unterricht in der Klasse" in: W. CORRELL (Hrsg.): "Programmiertes Lernen und Lehrmaschinen", Braunschweig 1966
- GLEASON, G.: "Programed learning in sixth grade arithmetic", Univ. of Wisconsin, Milwaukee 1964 (AERA-Paper)
- GOTKIN, I.G.: "Experimentation with Programed Instructions: Some Suggestions for Educators", in: Programed Instruction, 1,1, 1962
- GUTHRIE, E.R.: "The psychology of learning", New York 1952
- HARDE, O.: "Ein Jahrgangsprogramm für den Rechenunterricht", in: B. ROLLET (Hrsg.): "Praxis und Theorie des Programmierten Unterrichts", Stuttgart-München, 1970
- HEINRICHS, H.: "Roboter vor der Schultür", Bochum 1964
- HEINRICHS, H.: "Erste Ergebnisse des Schulfernsehversuchs Hannover", in: Deutsche Lehrprogramme für Schule und Praxis, 2, 1965
- HEIPKE, K.: "Die Bedeutung mathematischer Lernmodelle bei der Auswertung von Lernversuchen", in: pl, 3, 1968
- HILGARD, E.R.: "Kernprobleme der Lernpsychologie und des programmierten Lernens", in: Die Deutsche Schule, 1963
- HILGARD, E.R.: "Theorien des Lernens I", Stuttgart 1970

- HIRZEL, M.: "Partnerarbeit im Programmierten Unterricht", Stuttgart 1970
- HOCHHEIMER, W.: "Zur Tiefenpsychologie des pädagogischen Feldes", in: Die deutsche Schule, 51. Jg., 3, 1959
- HOCHHEIMER, W.: "Psychologische Einführung in die Thematik von Lehrmaschinen und Programmierter Unterricht", in: A. BEANCHERI (Hrsg.): "Programmierter Unterricht und Lehrmaschinen. Bericht Internationale Konferenz Berlin 1963", Berlin 1964
- HOCHHEIMER, W.: "Untersuchungen und Erprobungen zum Programmieren Lernen", in: ZeF (Zeitschrift für erziehungswissenschaftliche Forschung), 1, 1967, 1
- HOFSTÄTTER, P.R.: "Tiefenpsychologische Persönlichkeits-theorien", in: H. LERSCH u. H. THOMAE (Hrsg.): "Persönlichkeitsforschung und Persönlichkeits-theorie", Hb. d. Psych., Bd. 4, Göttingen 1960
- HOFSTÄTTER, P.R. u. WENDT, D.: "Quantitative Methoden der Psychologie", München 1967
- HOLLAND, J.G.: "Reaktionszusammenhänge in Lehrmaschinenprogrammen", in: W. CORRELL (Hrsg.): "Zur Theorie und Praxis des programmierten Lernens", Darmstadt 1969
- HOLLAND, J.G. and SKINNER, B.F.: "The Analysis of behavior", New York 1961
- HORN, W.: "Leistungsprüfsystem L-P-S", Göttingen 1962
- HUBER, G.: "Lernpsychologische Befunde bei Programmierter Unterweisung", München 1966

- INGENKAMP, K.-H. (Hrsg.): "Schulkonflikt und Schülerhilfe", Weinheim 1965
- INGENKAMP, K.-H.: "Das Testen kognitiver Fähigkeiten und Leistungen", in: K.-H. INGENKAMP (Hrsg.): "Handbuch der Unterrichtsforschung Teil I", Weinheim-Berlin-Basel, 1970
- INGLE, R. and GEPHARDT, W.J.: "A Critique of a Research Report: Programed Instruction Versus Usual Classroom Procedures in Teaching Boys to Read", in: Am. Educ. Research Journal, Bd. III, I, 1966
- ISSING, L.J.: "Forschung und Forschungsergebnisse im Bereich des Programmierten Unterrichts", in: L.J. ISSING (Hrsg.): "Der Programmierte Unterricht in den USA heute", Weinheim 1967
- ISSING, L.J. und SCHELLENBERG, C.: "Anwendung von PU-Prinzipien auf die Gestaltung von Schulfernsehsendungen - eine Vergleichsuntersuchung", in: pl, 1969
- JAYNE, D.C.: "A study of the relationship between teaching procedures and educational outcomes", in: Journ. Exp. Educ., 1945, 14
- KAY, H.: "General introduction to teaching machine procedures", in: K. AUSTWICK (Ed.): "Teaching Machines and programming", Oxford 1964
- KEIL, W. und SADER, M.: "Situationsspezifische Einflüsse auf die Leistungsmotivationsmessung nach Heckhausen", in: Ztschr. f. exp. u. ang. Psych., 1, 1968

- KEISLAR, E.: "The Development of Understanding in Arithmetic by a Teaching Machine", in: J.P. De CECCO (Hrsg.): "Educational Technology", New York 1964
- KLAUER, K.J.: "Über die Erfassung der Validität von Lehrprogrammen", in: pl (programmiertes lernen und programmierter unterricht), 1, 1968
- KLAUER, K.J.: "Orthographisches Lernen als Funktion der Lehrmethode, des Leistungs- und des Intelligenzniveaus", in: ZeF, 2, 1968, 1
- KLAUER, K.J.: "Die Bedeutung des Klasseneffekts für die Schulpädagogische Forschung", in: pl, 3, 1970
- KLAUER, K.J.: "Über Möglichkeiten der Experimentalforschung in Schulklassen", in: pl, 1, 1971
- KLAUS, D.J.: "Die Programmierung des Selbstunterrichts", in: W. CORRELL (Hrsg.): "Programmiertes Lernen und Lehrmaschinen", Braunschweig 1966
- KRANICH, E.M.: "Pädagogische Projekte und ihre Folgen", Stuttgart 1969
- KRUMBOLTZ, J.D. and WEISMAN, R.G.: "The effect of overt versus covert responding to programmed instruction on immediate and relayed retention", in: Journ. educ. Psychol., 1962, 53
- KRUMBOLTZ, J.D. and WEISMAN, R.G.: "The Effect of Intermittent Confirmation in Programmed Instruction", in: J.P. De CECCO (Hrsg.): "Educational Technology", New York 1964
- KULKARNI, S.S.: "Kriteriumstests und Programmiertes Lernen", in: K.-H. INGENKAMP und Th. MARSOLEK (Hrsg.): "Möglichkeiten und Grenzen der Testanwendung in der Schule", Weinheim-Berlin-Basel, 1968

- LAMBERT, P.; MILLER, D.M. and WILEY, D.E.: "Experimental folklore and experimentation: The study of programmed learning in the Wauwatosa public Schools", in: Journ. Educ. Res., 1962, 55, (9)
- LANDA, L.N.; ORLOWA, A.M.; GRANIK, G.G.: "Prinzipien des Programmierten Unterrichts in russischer Sprache", in: W. CORRELL (Hrsg.): "Zur Theorie und Praxis des programmierten Lernens", Darmstadt 1969
- LEITH, G.O.; DAVIS, M. and TERENCE, N.: "The influence of social reinforcement and personality on achievement in a school learning task", in: Research Report No. 19, National Center for Programmed Learning, Birmingham 1967
- LEONTJEW, A.N. und GALPERIN, P.J.: "Die Theorie des Kenntniserwerbs und der Programmierte Unterricht", in: W. CORRELL (Hrsg.): "Zur Theorie und Praxis des programmierten Lernens", Darmstadt 1969
- LIENERT G.A.: "Verteilungsfreie Methoden in der Biostatistik", Meisenheim 1962
- LIENERT, G.A.: "Eine varianzanalytische Methode zum Nachweis der Gruppenspezifität von Testprofilen", in: Zt. f. exp. u. angew. Psych., 10, 1963
- LINDQUIST, E.F.: "Designs and analysis of experiments in psychology and education", Boston 1953
- LINDVALL, G.M.: "Pupil attention and attitude under conditions of programed instruction", Univ. of Pittsburgh, 1964 (AERA-Paper)

LUMSDAINE, A.A.: "Assesing the effectiveness of instructional programs", in: R. GLASER (Ed.): "Teaching Machines and programed learning II", New York 1965

LUMSDAINE, A.A.: "Probleme beim programmierten Unterricht", in: W. CORRELL (Hrsg.): "Zur Theorie und Praxis des programmierten Lernens", Darmstadt 1969

LUSCHER, H.: "Programmierter Unterricht im Spiegel erster Erfahrungen", Braunschweig 1964

LYSAUGHT, J.P. und WILLIAMS, C.M.: "Einführung in die Unterrichtsprogrammierung", München-Wien 1967

MAGER, R.F.: "Preparing Objectives for Programmed Instruction", San Francisco 1962

MAGER, R.F.: "On the Sequencing of Instructional Content", in: J.P. De CECCO (Ed.): "Educational Technology", New York 1964

MECHNER, F.: "Programming for automated instruction", New York: Basic Systems, Inc., 1961

MEDLEY, D.M. and MITZEL, H.E.: "Measuring Classroom Behavior by Systematic Observation", in: N.L. GAGE (Ed.): "Handbook of Research on Teaching", Chicago 1965

MILLER, D.M.; WILIEY, D.E.; LAMBERT, P.: "An experimental study of the effect of instructional method on the acquisition, retention and transfer of English punctuation skills", Univ. of Wisconsin, 1964 (AERA-Paper)

- MITTENECKER, E.: "Planung und statistische Auswertung von Experimenten", Wien 1963
- MIZERA, A. und FISNER, B.: "Ein Versuch mit dem programmierten Unterricht der Chemie an der Maschinenbaufakultät der Tschechischen Technischen Hochschule in Prag", in: pl 2, 1968
- MÜLLER, L.: "Gaudigs freie geistige Schularbeit und das Programmierte Lernen", Schweizerische Lehrerzeitung, Zürich 111, 1966
- NOHL, H.; WENIGER, E.; GEISLER, G. (Hrsg.): "Das Problem der Unterrichtsmethode", Kleine Pädagogische Texte, Heft 19, Weinheim o.J.
- PAGE, E.B.: "Teacher comments and student performance: A seventyfour classroom experiment in school motivation", in: Journ. educ. Psychol., 1958, 49
- PRESSEY, S.L.: "Eine Maschine zum automatischen Lehren von Übungsstoffen", in: W. CORRELL (Hrsg.): "Programmiertes Lernen und Lehrmaschinen", Braunschweig 1966
- PRESSEY, S.L.: "Ein Gerät, das Tests darbietet, sie auswertet und lehrt", in: W. CORRELL (Hrsg.): "Programmiertes Lernen und Lehrmaschinen", Braunschweig 1966

- RATLIFF, F. and BLOUGH, D.S.: "Behavioral studies of visual processes in the pigeon", Report of Contract N5ori-07663, Psychological Laboratories, Harv. Univ., Sept. 1954
- ROE, A.: "Automated Teaching methods using linear programs", in: Journ. appl. Psychol., 1962, 46
- ROE, A.: "A Comparison of Branching Methods for Programmed Learning", in: Journ. of Educ. Res., 55, 1962
- ROE, A.; MASSEY, M.; WELTMAN, G. and LEEDS, D.: "Automated Teaching Methods using linear Programs", Report Nr. 60-105, December 1960. Department of Engineering, Univ. of California, Los Angeles
- RÖTTEL, K.: "Vergleichende Analyse einer Versuchsreihe mit programmiertem und konventionellem Unterricht", in: ZeF (Zeitschrift für erziehungswissenschaftliche Forschung) 2, 1968, 4
- ROTHKOPF, E.Z.: "The Epidemiology of Ignorance", in: Contemp. Psychol., 9, 1964
- SCHENK-DANZIGER, L.: "Schulreife und Erziehungsprobleme im Rahmen des Anfangsunterrichts", in: H. HETZER (Hrsg.): "Pädagogische Psychologie", Hb. d. Psych., Bd. 10, Göttingen 1959
- SCHIEFELE, H.: "Programmierte Unterweisung", München 1964
- SCHRAMM, W.: "Programmierter Unterricht heute und morgen", Berlin-Bielefeld, 1963

- SCHRAMM, W.: "Four Case Studies of Programed Instruction", New York 1964
- SCHRÖTER, G.: "Objektivierung des Unterrichts", Baunschweig 1965
- SCHRÖTER, G.: "Untersuchungen über die Einstellung von Berufsschülern zur Programmierten Unterweisung", in: "Praxis und Perspektiven des Programmierten Unterrichts", Quickborn 1965
- SCHÜTT, F.W.: "Unterrichtsversuche im Chemieunterricht des Gymnasiums bei programmiertem und konventionellem Unterricht", in: pl, 3, 1968
- SCHULZ, W.; TESCHNER, W.P. und VOIGT, J.: "Verhalten im Unterricht. Seine Erfassung durch Beobachtungsverfahren", in: K.-H. INGENKAMP (Hrsg.): "Handbuch der Unterrichtsforschung Teil I", Weinheim-Berlin-Basel, 1970
- SCHWARZ, E.: "Experimentelle und quasiexperimentelle Anordnungen in der Unterrichtsforschung", in: K.-H. INGENKAMP (Hrsg.): "Handbuch der Unterrichtsforschung Teil I", Weinheim-Berlin-Basel, 1970
- SILBERMAN, H.E.: "Trends in der Erziehungstechnologie", in: L.J. ISSING (Hrsg.): "Der Programmierte Unterricht in den USA heute", Weinheim 1967
- SHAY, C.B.: "Relationship of Intelligence to Step Size on a Teaching Machine Program", in: Journ. of Educ. Psychol., 52, 1961
- SKINNER, B.F.: "The behavior of organism: an experimental analysis", New York 1938

SKINNER, B.F.: "Science and human behavior", New York
1953

SKINNER, B.F.: "Lehrmaschinen", in: W. CORRELL (Hrsg.):
"Programmiertes Lernen und Lehrmaschinen",
Braunschweig 1966

SKINNER, B.F.: "Die Wissenschaft vom Lernen und die
Kunst des Lehrens", in: W. CORRELL (Hrsg.):
"Programmiertes Lernen und Lehrmaschinen",
Braunschweig 1966

SKINNER, B.F.: "Fallstudie zur wissenschaftlichen Methode
in der Psychologie", in: W. CORRELL (Hrsg.):
"Programmiertes Lernen und Lehrmaschinen",
Braunschweig 1966

SKINNER, B.F.: "Warum wir Lehrmaschinen brauchen", in:
W. CORRELL (Hrsg.): "Programmiertes Lernen
und Lehrmaschinen", Braunschweig 1966

SKINNER, B.F.: "Verhaltenspsychologische Analyse des
Denkprozesses", in: B.F. SKINNER und W.
CORRELL: "Denken und Lernen", Braunschweig 1967

SKINNER, B.F.: "Warum Lehrer versagen", in: W. CORRELL
(Hrsg.): "Zur Theorie und Praxis des pro-
grammierten Lernens", Darmstadt 1969

SKINNER, B.F.: "Die Technologie des Lehrens", in: W.
CORRELL (Hrsg.): "Zur Theorie und Praxis
des programmierten Lernens", Darmstadt 1969

SKOWRONEK, H.: "Psychologische Grundlagen einer Didaktik
der Denkerziehung", Hannover 1968

STÖCKER, K.: "Neuzeitliche Unterrichtsgestaltung", Mün-
chen 1970

- STOLUROW, L.M.: "Teaching by Machine", Washington 1961
- STOLUROW, L.M.: "A Response to 'inside Opinion'", in:
Programed Instruction, 1, 2, 1962
- STOLUROW, L.M. and LIPPERT, H.: "Prompting, Confirmation
and Overlearning in the Automated Teaching of
a Sight Vocabulary", in: J.P. De GECCO (Ed.):
"Educational Technology", New York 1964
- STOLUROW, L.M. and WALKER, C.C.: "A comparison of overt
and covert response in programmed learning",
in: Journ. educ. res., 1962, 55
- STRITTMATTER, P.: "Präsentationsmodi von Lehrprogram-
men", Weinheim 1970
- SÜLLWOLD, F.: "Zur Frage der Präzision und Ergiebig-
keit von pädagogischen Experimenten", in:
"Pädagogische Forschung und Pädagog. Praxis",
Veröffentl. d. Hochsch. f. intern. päd. Forschg.,
1, 1958
- SÜLLWOLD, F.: "Problemfragebogen für Jugendliche", Göttin-
gen 1968
- TENT, L.: "Schätzverfahren in der Unterrichtsforschung",
in: K.-H. INGENKAMP (Hrsg.): "Handbuch der
Unterrichtsforschung Teil I", Weinheim-Ber-
lin-Basel 1970
- THOMAE, H.: "Das Individuum und seine Welt. Eine Per-
sönlichkeitstheorie", Göttingen 1968
- THORNDIKE, E.L.: "Psychologie der Erziehung", Darmstadt
1970

- TOMAN, W.: "Motivation, Persönlichkeit, Umwelt", Göttingen 1968
- VOIGT, H.: "Pädagogische Kybernetik und Programmierte Instruktion in der Sowjetunion und in der DDR 1965/66 mit Bibliographie", Weinheim 1968
- WALTER, H.: "Das Problem der Validierung von Unterrichtsprogrammen", in: G. DIETRICH (Hrsg.): "Beiträge zur Unterrichtspsychologie der Hauptschule", erscheint München 1971
- WALLIS, D. and WICKS, R.P.: "The Autotutor and Classroom Instruction: Three Comparative Studies. 1. The Royal Navy Study", in: Occupational Psychology, 1963, 37
- WENDT, P.R. and RUST, G.: "Pictorial and Performance Frames in Branching Programed Instruction", in: Journ. of Educ. Res., 55, 1962
- WERTHEIMER, M.: "Produktives Denken", Frankfurt 1957
- WILLIAMS, J.P.: "Effectiveness and Constructed-Response and Multiple-Choice Programming Modes as a Function of Test Mode", in: Journ. of Educ. Psychol. 56, 1965
- WINDLE, C.: "Test-retest effect on personality questionnaires", in: Educ. psychol. Measmt., 1954, 14

WINNEFELD, F.W.: "Pädagogischer Kontakt und Pädagogisches Feld", München-Basel, 1957

WITTE, A.: "Experimentelle Beiträge zur Entwicklung des kombinierten Unterrichts", in: "Lehrmaschinen in kybernetischer und pädagogischer Sicht", Stuttgart-München 1965

WOODWORTH, R.S.: "Experimental Psychology" New York 1954

ZIFREUND, W. (Hrsg.): "Schulmodelle, Programmierte Instruktion und Technische Medien", München 1968

ZIFREUND, W.: "Über den Zusammenhang von Programmierter Instruktion, Unterrichtstechnologie und Unterrichtsforschung", in: pl, 1, 1969

ZILLIG, M.: "Psychologische Probleme der Oberstufe der Volksschule", in: H. HETZER (Hrsg.): "Pädagogische Psychologie", Hb. d. Psych., Bd. 10, Göttingen 1959

ANHANG

Schüler- Nummer	Vorwissen	KNO / PU	COM / PU	KNO / KU	COM / KU	Schüler Nummer
1	4	15,25	3,75	13,00	2,75	1
2	8	17,75	8,00	17,50	10,00	2
3	3	14,75	7,00	11,00	2,50	3
4	4	14,00	6,25	13,50	2,00	4
5	4	15,50	4,00	14,25	8,25	5
6	9	16,75	3,50	10,75	2,50	6
7	6	17,50	2,25	12,00	2,75	7
8	8	17,25	2,00	14,25	4,25	8
9	3	14,25	2,50	9,50	3,50	9
10	4	18,25	8,25	10,00	3,25	10
11	5	18,00	7,25	11,00	2,00	11
12	7	15,00	1,75	13,50	3,00	12
13	9	9,75	3,50	15,50	2,25	13
14	3	17,50	5,00	15,50	5,75	14
15	5	17,50	1,00	12,00	4,50	15
16	5	13,75	5,00	12,25	6,00	16
17	6	11,75	2,50	11,50	6,50	17
18	4	8,75	2,00	12,50	1,50	18
19	6	15,00	2,75	13,00	3,00	19
20	4	13,00	2,25	10,50	1,25	20
21	6	18,00	7,75	18,00	3,75	21
22	6	12,50	4,50	14,50	2,00	22
23	6	9,75	4,00	9,75	1,00	23
24	2	10,00	3,00	13,75	1,00	24
25	7	15,50	3,50	11,50	6,75	25
26	6	10,00	1,50	12,50	1,00	26
27	7	19,00	6,00	18,00	2,25	27
28	7	10,50	7,00	16,50	3,00	28
29	5	14,00	2,25	18,00	4,00	29
30	7	14,00	3,00	13,25	2,50	30
31	6	15,00	4,50	18,00	5,50	31

Schüler Nummer	Vorwissen	KNO / PU	COM / PU	KNO / KU	COM / KU	Schüler Nummer
32	6	15,00	5,00	9,75	1,00	32
33	5	16,00	2,00	13,50	1,50	33
34	6	16,00	7,25	14,00	3,50	34
35	5	10,50	2,25	15,00	0,50	35
36	4	17,00	2,00	16,00	0,75	36
37	9	16,75	10,25	15,25	5,50	37
38	5	13,00	3,00	13,75	3,25	38
39	6	15,50	2,75	10,50	4,50	39
40	7	12,00	6,25	11,50	0,25	40
41	4	8,00	3,00	10,50	2,25	41
42	5	17,75	4,50	13,00	3,50	42
43	6	6,00	1,25	6,00	0,75	43
44	7	14,75	3,00	15,00	4,00	44
45	7	19,75	8,25	15,00	3,50	45
46	4	15,00	5,00	15,00	3,00	46
47	6	11,75	2,00	9,50	3,75	47
48	9	16,25	9,00	13,50	6,00	48
49	7	16,50	4,25	17,75	5,00	49
50	8	15,25	3,75	15,25	4,50	50
51	6	12,75	5,00	16,50	3,00	51
52	10	16,50	8,00	16,50	4,00	52
53	4	8,00	2,00	11,00	3,00	53
54	7	17,00	3,25	18,00	5,75	54
55	7	16,25	4,00	16,00	6,50	55
56	5	16,00	4,00	13,50	2,00	56
57	6	16,75	2,00	16,00	4,00	57
58	8	18,00	3,00	20,00	6,00	58
59	9	17,25	1,00	15,00	3,00	59
60	9	15,75	3,00	10,00	2,50	60
61	6	13,00	1,00	16,00	2,25	61
62	10	20,00	10,25	19,50	10,25	62

Schüler- Nummer	SW gesamt	SW math.	Einstellg. PU	Einstellg. KU	Einst. z. Lehrer	Schüler Nummer
1	118	106	16	11	22	1
2	127	119	20	16	26	2
3	128	122	17	15	23	3
4	108	102	16	17	31	4
5	116	108	21	16	21	5
6	110	104	14	15	20	6
7	110	107	13	14	26	7
8	124	109	19	15	22	8
9	113	115	10	15	24	9
10	110	94	18	18	21	10
11	105	99	20	12	19	11
12	116	100	23	11	19	12
13	112	98	23	12	20	13
14	120	110	19	12	18	14
15	115	106	19	18	27	15
16	113	107	22	15	19	16
17	107	91	14	18	21	17
18	97	91	23	13	24	18
19	108	99	16	11	22	19
20	110	106	20	16	26	20
21	141	128	15	11	19	21
22	118	110	20	6	19	22
23	107	92	19	12	12	23
24	109	99	23	20	22	24
25	111	100	22	8	13	25
26	112	101	30	14	12	26
27	120	109	18	6	12	27
28	111	104	30	12	11	28
29	126	124	22	10	20	29
30	117	107	26	9	20	30
31	127	115	14	9	19	31

Schüler- Nummer	SW gesamt	SW math.	Einstellg. PU	Einstellg. KU	Einst. z. Lehrer	Schüler Nummer
32	117	107	22	17	13	32
33	123	114	18	15	21	33
34	118	112	25	11	19	34
35	125	109	23	10	14	35
36	125	110	13	13	16	36
37	110	109	13	9	11	37
38	109	100	30	11	12	38
39	117	100	24	21	24	39
40	111	106	15	10	18	40
41	118	114	25	10	13	41
42	108	107	28	9	10	42
43	107	106	24	11	21	43
44	124	106	17	12	15	44
45	125	107	18	9	15	45
46	127	117	19	15	21	46
47	111	98	18	10	15	47
48	129	106	25	14	15	48
49	118	112	19	19	16	49
50	115	115	29	9	22	50
51	110	97	16	20	22	51
52	130	113	10	13	20	52
53	108	93	21	17	18	53
54	114	111	21	16	19	54
55	122	122	17	12	13	55
56	112	101	22	16	18	56
57	121	104	18	24	30	57
58	134	111	16	21	20	58
59	129	117	15	25	21	59
60	114	112	28	18	26	60
61	134	116	16	20	29	61
62	128	121	15	14	16	62

Schüler- Nummer	Arbeitsh. "Sozial"	Arbeitsh. "Konzentr."	Einst. z. Schule	Verstär- kungen	Ant- worten	Schüler Nummer
1	9	7	45	10	19	1
2	8	8	67	8	13	2
3	8	10	32	24	30	3
4	9	9	42	8	11	4
5	9	7	54	8	21	5
6	10	9	62	3	5	6
7	8	10	46	10	16	7
8	9	9	43	11	14	8
9	6	8	60	21	25	9
10	8	8	39	2	8	10
11	9	9	46	2	4	11
12	10	9	52	4	5	12
13	9	9	60	3	5	13
14	7	12	61	19	24	14
15	9	10	42	11	13	15
16	8	9	62	10	16	16
17	10	8	62	8	12	17
18	8	3	31	8	11	18
19	10	9	47	9	14	19
20	7	6	63	8	16	20
21	14	14	43	12	14	21
22	12	11	54	13	13	22
23	9	7	34	2	3	23
24	9	6	73	8	14	24
25	7	9	43	15	16	25
26	8	6	66	12	17	26
27	10	12	50	11	15	27
28	10	12	30	22	31	28
29	13	12	72	4	5	29
30	9	10	40	11	19	30
31	10	11	63	18	19	31

Schüler- Nummer	Arbeitsh. "Sozial"	Arbeitsh. "Konzentr."	Einst. z. Schule	Verstär- kungen	Ant- worten	Schüler- Nummer
32	11	10	30	7	14	32
33	11	10	44	10	11	33
34	12	8	45	13	16	34
35	9	7	44	16	17	35
36	10	10	53	13	22	36
37	12	14	57	13	14	37
38	5	6	40	16	28	38
39	11	10	49	8	9	39
40	10	9	35	9	15	40
41	10	10	47	14	15	41
42	11	9	51	10	10	42
43	9	6	66	8	9	43
44	10	8	35	11	13	44
45	11	14	40	16	18	45
46	10	8	37	10	12	46
47	9	7	47	11	14	47
48	12	15	74	12	15	48
49	7	12	63	27	41	49
50	9	11	53	5	7	50
51	6	9	48	12	20	51
52	11	11	56	13	17	52
53	7	9	43	7	12	53
54	7	9	55	6	9	54
55	5	9	44	15	19	55
56	11	10	34	10	12	56
57	11	11	57	9	10	57
58	6	11	62	12	18	58
59	10	8	39	9	15	59
60	11	9	44	3	15	60
61	8	8	52	12	20	61
62	14	15	56	13	17	62

Schüler- Nummer	Vorwissen	KNO / PL	COM / PU	KNO / KU	COM / KU	Schüler Nummer
63	6	15,00	5,75	19,00	5,75	63
64	7	14,00	2,50	11,50	0,00	64
65	7	17,50	6,00	19,00	4,25	65
66	9	18,00	3,75	17,75	3,25	66
67	6	16,50	3,50	14,00	1,00	67
68	9	13,00	0,00	15,50	1,00	68
69	6	17,50	3,25	16,75	2,00	69
70	4	14,75	2,50	13,00	5,00	70
71	4	8,50	1,50	11,25	2,00	71
72	6	16,00	3,50	13,75	2,25	72
73	4	10,50	0,25	9,00	1,00	73
74	5	17,00	7,50	16,00	2,25	74
75	8	12,50	4,00	17,50	6,50	75
76	3	11,50	0,00	8,00	2,00	76
77	7	14,50	2,75	19,00	1,00	77
78	8	15,50	0,00	12,75	2,25	78
79	5	12,50	1,00	16,25	0,00	79
80	11	19,50	14,75	20,00	10,25	80
81	6	12,00	3,50	16,00	3,25	81
82	6	16,00	7,25	17,50	3,00	82
83	9	13,50	3,00	7,50	1,50	83
84	6	15,50	4,25	20,00	7,25	84
85	4	12,50	4,00	17,75	3,25	85
86	0	9,50	7,00	14,50	4,00	86
87	7	15,25	5,50	16,50	2,50	87
88	10	20,00	9,00	16,25	9,50	88
89	9	15,50	6,00	14,00	5,25	89
90	8	13,50	5,00	15,00	4,00	90
91	9	17,00	10,25	19,50	5,00	91
92	6	15,50	2,00	8,75	1,25	92
93	9	19,50	9,75	18,00	8,00	93

Schüler Nummer	Vorwissen	KNO / PL	COM / PL	KNO / KU	COM / KU	Schüler Nummer
94	7	18,00	10,25	18,00	4,00	94
95	8	17,50	6,25	15,50	3,00	95
96	8	16,50	5,00	17,00	3,00	96
97	5	17,50	6,75	13,50	4,00	97
98	4	19,50	9,75	17,50	4,00	98
99	0	18,50	7,25	18,75	4,25	99
100	9	18,00	11,00	14,00	1,25	100
101	5	17,00	1,25	12,00	1,00	101
102	4	11,00	0,75	10,25	3,25	102
103	4	19,00	2,00	10,25	0,00	103
104	5	15,50	3,50	9,50	5,50	104
105	3	11,50	0,50	5,75	0,50	105
106	4	12,50	3,50	8,00	0,00	106
107	5	15,00	4,25	17,00	5,00	107
108	2	10,50	0,00	8,25	1,25	108
109	5	13,50	1,50	10,75	3,50	109
110	3	14,00	0,50	14,00	2,00	110
111	5	16,50	1,25	6,75	0,00	111
112	6	15,50	2,00	11,25	0,00	112
113	5	17,00	1,25	12,00	1,50	113
114	6	15,50	1,50	13,00	3,25	114
115	5	15,50	4,00	14,00	4,25	115
116	6	13,00	2,00	12,75	4,00	116
117	4	19,25	1,50	13,75	4,00	117
118	4	16,50	2,25	16,00	4,00	118
119	6	16,00	2,75	10,75	2,00	119

Schüler- Nummer	SW gesamt	SW math.	Einstellg. PU	Einstellg. KU	Einst. z. Lehrer	Schüler Nummer
63	110	108	30	14	18	63
64	112	110	18	16	18	64
65	102	94	30	16	19	65
66	119	106	14	9	10	66
67	112	103	13	16	16	67
68	119	101	17	11	13	68
69	106	100	25	15	19	69
70	105	96	18	10	14	70
71	112	104	19	9	16	71
72	127	116	12	13	9	72
73	110	102	17	23	14	73
74	135	116	20	18	21	74
75	118	107	14	13	17	75
76	105	96	26	25	24	76
77	121	106	12	8	15	77
78	112	98	16	16	18	78
79	114	109	6	24	16	79
80	131	126	14	9	11	80
81	125	108	15	15	18	81
82	126	116	24	15	20	82
83	121	110	18	12	12	83
84	116	98	19	15	20	84
85	118	104	27	12	17	85
86	108	99	20	18	15	86
87	132	107	17	10	14	87
88	117	106	17	15	28	88
89	103	107	15	19	18	89
90	107	104	25	19	37	90
91	120	112	21	24	34	91
92	98	99	21	15	28	92
93	138	129	16	11	20	93

Schüler Nummer	Arbeitsh. "Sozial"	Arbeitsh. "Konzentr."	Einst. z. "Schule"	Verstär- kungen	Ant- worten	Schüler Nummer
63	11	3	44	13	17	63
64	13	8	45	2	3	64
65	13	9	50	8	9	65
66	10	12	33	31	41	66
67	9	14	49	22	41	67
68	12	13	82	2	4	68
69	12	12	72	10	15	69
70	10	8	46	23	30	70
71	11	10	82	20	26	71
72	12	10	37	11	15	72
73	13	8	52	8	12	73
74	5	11	78	17	26	74
75	10	10	54	3	4	75
76	12	4	57	10	20	76
77	7	12	34	16	21	77
78	12	8	68	5	6	78
79	13	7	62	3	4	79
80	14	15	64	2	8	80
81	7	6	69	9	12	81
82	11	12	40	11	15	82
83	7	6	32	17	23	83
84	12	15	46	19	27	84
85	10	9	48	7	13	85
86	11	10	54	13	15	86
87	12	12	45	3	6	87
88	10	10	54	4	5	88
89	10	6	62	14	17	89
90	9	8	40	8	9	90
91	12	10	48	3	5	91
92	8	7	57	13	12	92
93	9	12	36	14	14	93

Schüler Nummer	Arbeitsh. "Sozial"	Arbeitsh. "Konzentr."	Einst. z. Schule	Verstär- kungen	Ant- worten	Schüler Nummer
94	11	13	30	9	11	94
95	7	5	54	15	20	95
96	10	10	34	6	9	96
97	9	5	35	11	15	97
98	8	9	43	11	14	98
99	10	6	57	11	14	99
100	9	6	88	13	20	100
101	9	10	42	22	28	101
102	7	9	52	28	36	102
103	10	8	42	20	30	103
104	13	10	62	17	19	104
105	10	7	56	28	32	105
106	9	6	81	12	16	106
107	10	11	54	50	59	107
108	9	5	36	14	16	108
109	11	8	35	18	28	109
110	10	9	53	25	34	110
111	13	8	24	39	50	111
112	9	9	50	27	38	112
113	7	8	46	20	27	113
114	11	7	56	28	35	114
115	11	8	55	19	29	115
116	12	10	45	19	26	116
117	10	12	31	23	26	117
118	10	6	45	19	23	118
119	9	12	59	18	22	119

Prüfungsaufgaben

Wurzelrechnen

Nr.	p	Aufgabe	Lösung	Punktw.
1	0,99	Schreibe mit gebrochenem Exponent: $\sqrt[4]{b^5}$	$b^{5/4}$	1
2	0,90	Schreibe mit gebrochenem Exponent: $\sqrt[5]{(x+y)^2}$	$(x+y)^{2/5}$	1
3	0,88	Schreibe als Wurzel u. berechne gegebenenfalls: $(p+q)^{1/6}$	$\sqrt[6]{p+q}$	1
4	0,89	Schreibe als Wurzel u. berechne gegebenenfalls: $8^{2/3}$	4	1
5	0,73	Schreibe als Wurzel u. berechne gegebenenfalls: $4^{1,5}$	8	1
6	0,74	Berechne die Wurzel u. mache die Probe durch Potenzieren: $\sqrt[3]{125}$	5	1
7	0,57	Ziehe schriftlich die Quadratwurzel aus: $\sqrt{324}$	18	1
8	0,86	Schreibe als Wurzel: $y^{-1/4}$	$\frac{1}{\sqrt[4]{y}}$	1
9	0,61	Schreibe als Wurzel: $3^{-0,8}$	$\frac{1}{\sqrt[5]{3^4}}$	
10	0,63	Schreibe als Wurzel: $\left(\frac{p}{q}\right)^{-1/2}$	$\sqrt{\frac{q}{p}}$	

Nr.	P	Aufgabe	Lösung	Punktw.
11	0,83	Schreibe mit gebrochener Hochzahl: $\sqrt[3]{\frac{1}{m}}$	$m^{-1/3}$	1
12	0,69	Schreibe mit gebrochener Hochzahl: $\frac{1}{\sqrt[q]{q^y}}$	$q^{-\frac{y}{2}}$	1
13	0,50	Berechne mit Hilfe von gebrochenen Exponenten: $\frac{\sqrt[5]{a}}{\sqrt[3]{a^2}}$	$\sqrt[15]{a^7}$	1
14	0,85	Fasse soweit als möglich zusammen: $5^4\sqrt{q} + 3^4\sqrt{q} + 2^3\sqrt{q}$	$8^4\sqrt{q} + 2^3\sqrt{q}$	1
15	0,86	Multipliziere: $\sqrt[3]{a} \cdot \sqrt[3]{b} \cdot \sqrt[3]{c}$	$\sqrt[3]{abc}$	1
16	0,66	Löse durch Zerlegung: $\sqrt{4000}$	$20\sqrt{10}$	1
17	0,50	Dividiere und löse soweit als möglich: $\sqrt{64} : \sqrt{6}$	$\frac{4}{3}\sqrt{6}$	1
18	0,52	Löse soweit als möglich: $\sqrt[3]{0,064}$	0,4	1
19	0,51	Schreibe als Potenz mit rationaler Hochzahl: $\frac{1}{q\sqrt[q]{q+1}}$	$v^{-\frac{(q+1)}{q}}$	1
20	0,80	Vereinfache soweit als möglich: $\sqrt{\frac{p}{q}} \cdot \sqrt{\frac{q}{z}} \cdot \sqrt{\frac{z}{p}}$	1	1

Nr.	p	Aufgabe	Lösung	Punktw.
21	0,92	Welchen Wert hat die Wurzel: $\sqrt[14]{0}$	0	1
22	0,73	Vereinfache soweit als möglich: $\left(\sqrt[q]{b^p}\right)^{qm}$	b^{pm}	1
23	0,12	Vereinfache soweit als möglich: $\frac{(a-b)}{\sqrt{a}-\sqrt{b}}$	$\sqrt{a} + \sqrt{b}$	1
24	0,24	Multipliziere, $\left(\sqrt[7]{p^2}-\sqrt[7]{pq}+\sqrt[7]{p^2}\right)\left(\sqrt[7]{p}+\sqrt[7]{q}\right)$	$\sqrt[7]{p^2q}-\sqrt[7]{pq^2}+2\sqrt[7]{p^3}$	1
25	0,60	Löse: $\sqrt[5]{16^5}$	1024	1
26	0,84	Vereinfache: $\sqrt[12]{5^8}$	$\sqrt[3]{5^2}$	1
27	0,90	Vereinfache: $\sqrt[3]{\sqrt[4]{64}}$	2	1
28	0,24	Fasse soweit als möglich zusammen: $\frac{12}{a}\sqrt[3]{t-z} - \frac{3}{a}\sqrt[3]{t-q}$	$\frac{3}{a}\left(4\sqrt[3]{t-z} - \sqrt[3]{t-q}\right)$	1
29	0,13	Vereinfache: $2\sqrt[5]{49+4} - 4\sqrt[5]{9q+9} + 3\sqrt[5]{q+1}$	$\sqrt[5]{q+1}\left(2\sqrt[5]{4}-4\sqrt[5]{9+3}\right)$	1
30	0,18	Forme so um, daß keine Wurzeln im Nenner stehen: $\frac{p\sqrt{q} + q\sqrt{p}}{\sqrt{pq}}$	$\sqrt{p} + \sqrt{q}$	1

Nr.	P	Aufgabe	Lösung	Punktw.
31	0,57	Ziehe schriftlich die Wurzel: $\sqrt{409600}$	640	1
32	0,84	Dividiere und vereinfache soweit als möglich: $\sqrt[3]{\frac{125}{216}}$	$\frac{5}{6}$	1
33	0,69	Dividiere und vereinfache soweit als möglich: $\frac{\sqrt[4]{m^2n}}{\sqrt[4]{mn}}$	$\sqrt[4]{m}$	1
34	0,22	Vereinfache soweit als möglich: $\sqrt[3]{10p^4} - \sqrt[3]{q^4p^4} - \sqrt[3]{q^7p^{13}}$	$\sqrt[3]{qp}(q^3p - qp - q^2p^4)$	1
35	0,58	Drücke als eine Wurzel aus: $(m+n)\sqrt{\frac{m-n}{m+n}}$	$\sqrt{m^2 - n^2}$	1
36	0,68	Vereinfache soweit als möglich: $3\sqrt[3]{\frac{2}{3}}$	$\sqrt[6]{6}$	1
37	0,39	Ziehe soweit als möglich die Wurzel: $\sqrt{18q^3p^6 - 45q^4p^5}$	$3qp^2\sqrt{2p^2q - 5pq^2}$	1
38	0,30	Fasse soweit als möglich zusammen: $\left(\frac{2}{5}\sqrt{m} - \frac{3}{4}\sqrt{2}\right) - \left[\frac{2}{3}\sqrt{m} - \left(\frac{3}{4}\sqrt{2} + \frac{1}{2}\sqrt{m}\right)\right] + (-\sqrt{m})$	$\frac{37}{30}\sqrt{m}$	1
39	0,27	Vereinfache soweit als möglich: $\frac{2a}{\sqrt{m^2y^2 - a^2}} \cdot \frac{2a}{\sqrt{m^3a^2y^2 - 5a}}$	m^2y^2	1
40	0,39	Vereinfache soweit als möglich: $\frac{4}{\sqrt{27x^5}} \cdot \frac{4}{\sqrt{3x^3}}$	$3x^2$	1

Nr.	P	Aufgabe	Lösung	Punktw.
41	0,47	Schreibe als Wurzel: $a \cdot a^{5/6}$	$a^6 \sqrt{a^5}$	1
42	0,55	Vereinfache: $\sqrt[3]{8p} \cdot \sqrt[3]{2p^5}$	$4p^3$	1
43	0,42	Vereinfache: $\sqrt{0,2x} \cdot \sqrt{0,05x}$	$0,1x$	1
44	0,00	Vereinfache soweit als möglich: $\frac{\sqrt[3]{25x^5y^8}}{54p^8q^2} \cdot \sqrt[3]{\frac{10pq^5}{xy^4}}$	$\frac{xq^2}{3p^2} \sqrt[6]{\frac{25^7 \cdot xy^4}{pq}}$	1
45	0,14	Vereinfache soweit als möglich: $\frac{x-2\sqrt{xy}+y}{\sqrt{x}-\sqrt{y}}$	$\sqrt{x} - \sqrt{y}$	1
46	0,19	Prüfe die Gleichung: $\sqrt[3]{\frac{27}{8}} = 3\sqrt[3]{\frac{3}{8}}$	$\frac{\sqrt[3]{27}}{\sqrt[3]{8}} = 3 \frac{\sqrt[3]{3}}{\sqrt[3]{8}}$	1
47	0,15	Vereinfache soweit als möglich: $\sqrt[3]{\sqrt{x}}$	$\sqrt[6]{x}$	1
48	0,13	Vereinfache soweit als möglich: $\sqrt[5]{25(x+y)^4}$	$\frac{10}{\sqrt[5]{25}} \sqrt[5]{(x+y)^4}$	1
49	0,60	Vereinfache soweit als möglich: $\sqrt{3\sqrt{3} \cdot \sqrt{3}}$	3	1
50	0,15	Vereinfache soweit als möglich: $\sqrt[4]{\frac{a^3}{\sqrt{a}}}$	$\sqrt[8]{a^5}$	1

Prüfungsaufgaben

Logarithmen

Nr.	r	Aufgabe	Lösung	Punktw.
1	0,83	Schreibe in Potenzform und bestimme den Numerus: $\log_4 x = 3$	$4^3 = x$ $x = 64$	1
2	0,88	Schreibe in Potenzform und bestimme den Numerus: $\log_a y = b$	$a^b = y$	1
3	0,92	Berechne: $\log_2 16$	4	1
4	0,76	Berechne: $\log_5 \frac{1}{625}$	- 4	1
5	0,89	Berechne: $\log_5 5$	$\frac{1}{2}$	1
6	0,62	Berechne: $\log_a \frac{1}{4\sqrt[4]{a^4}}$	- 1	1
7	0,68	Berechne: $\log_{1/4} 0,25$	1	1
8	0,87	Schreibe in logarithmi- scher Form und bestimme den Exponenten: $2^x = 32$	$x = \log_2 32$ $x = 5$	1
9	0,70	Schreibe in logarithmi- scher Form und bestimme den Exponenten: $\sqrt{x} \sqrt{1000} = 10$	$x = 3$	1
10	0,66	Schreibe in logarithmi- scher Form und bestimme den Exponenten: $4^x = \frac{1}{64}$	$x = \log_4 \frac{1}{64}$	1

Nr.	p	Aufgabe	Lösung	Punktw.
11	0,85	Zerlege in Summen und berechne: $\log_{10}(10 \cdot 1000 \cdot 10\,000\,000)$	$\log_{10} 10 + \log_{10} 1000 +$ $\log_{10} 10\,000\,000$	1
12	0,86	Zerlege in Differenzen und berechne: $\log_2 \frac{32}{128}$	$\log_2 32 - \log_2 128$ $= -2$	1
13	0,81	Fasse zusammen und berechne: $\log_3 189 - \log_3 7$	$\log_3 27$ $x = 3$	1
14	0,80	Fasse zusammen und berechne: $\log_2 3 + \log_2 \frac{2}{3} + \log_2 4$	$\log_2 8$	1
15	0,68	Fasse zusammen und berechne: $\log_n(q^2 - p^2) - \log_n(q + p)$	$\log_n(q - p)$	1
16	0,82	Bringe in Potenzform und berechne dann: $\log_x 8 = \frac{1}{2}$	$x = 64$	1
17	0,45	Bringe in Potenzform und berechne: $\log_x \frac{1}{2} = -\frac{1}{3}$	$x = 8$	1
18	0,68	Bringe in Potenzform und berechne: $\log_x 144 = 2$	$x = 12$	1
19	0,73	Stelle als Produkt dar: $\log_a \sqrt[q]{p}$	$\frac{1}{q} \log_a p$	1
20	0,73	Stelle als Produkt dar: $\log_b 10^4$	$4 \cdot \log_b 10$	1

Nr.	p	Aufgabe	Lösung	Punktw.
21	0,93	Stelle als Produkt dar: $\log_a p^q$	$q \log_a p$	1
22	0,69	Stelle als Produkt dar: $\log_a \sqrt[4]{10}$	$\frac{1}{4} \log_a 10$	1
23	0,53	Zerlege soweit als möglich: $\log_a \frac{q \cdot \sqrt[m]{q}}{a^2 \cdot z^3}$	$\log_a q + \frac{1}{m} \log_a q =$ $2 \log_a a + 3 \log_a z$	1
24	0,30	Schreibe als Logarithmus einer Zahl: $3 + \log_a 5$	$\log_a 5a^3$	1
25	0,25	Schreibe als Logarithmus einer Zahl: $4 - \log_a 2$	$\log_a \frac{a^4}{2}$	1
26	0,12	Berechne y: $\log y = \frac{1}{2} \log(u^2 - v^2) - \log(u+v)$	$y = \sqrt{\frac{u-v}{u+v}}$	1
27	0,21	Berechne y: $\log y = 2 \log(u-v) + \frac{1}{2} \log(u+v)$	$y = (u-v)^2 \cdot \sqrt{u+v}$	1
28	0,10	Berechne y: $\frac{1}{2} \log y = 2 \log(u+v) + \frac{1}{2} \log u$	$y = u \cdot (u+v)^4$	1
29	0,22	Rechne soweit als möglich: $\log_{1/8} 0,25$	$\frac{2}{3}$	1
30	0,70	Rechne soweit als möglich: $\log_{\sqrt{5}} 25$	4	1

Nr.	p	Aufgabe	Lösung	Punktw.
31	0,58	Begründe: $\log(u-v) \neq \log u - \log v$	$\log u - \log v = \log \frac{u}{v}$	1
32	0,25	Zerlege soweit als möglich: $\log \sqrt[3]{\frac{u^2 v^3}{u+v}} \cdot 0,5^{1/2}$	$\frac{1}{3} \left[2 \log u + 3 \log v - \frac{1}{2} \log 0,5 - \log(u+v) \right]$	1
33	0,11	Berechne soweit als möglich: $\log \sqrt{m} \cdot m \cdot \sqrt{m} = x$	$x = \frac{7}{2}$	1
34	0,08	Berechne soweit als möglich: $\log(x-1) - \log x = 3 \log m - 2 \log mn^{1/2}$	$x = \frac{n}{n-m^2}$	1
35	0,05	Logarithmiere: $\frac{\left(\frac{1}{u} + \frac{1}{v}\right)(u-v)}{\left(\frac{1}{u} - \frac{1}{v}\right)(u+v)}$	$\log_{10}(u-v) - \log_{10}(v-u)$	1
36	0,02	Logarithmiere: $\frac{\sqrt{x^3} \sqrt{x}}{\sqrt[3]{x} \sqrt{x}}$	$\frac{5}{4} \log_{10} x$	1
37	0,01	Logarithmiere: $\frac{\sqrt{1-u^2} (2u+5)}{\frac{1}{\sqrt{1-u^2}} (3u-1)}$	$\log(2u+5) + \log(1-u^2) - \log(3u-1)$	1
38	0,01	Entlogarithmiere: $\log \left(\frac{7u}{2v^2} - 1 \right)^2 - \log \left(\frac{2v^2}{7u} + 1 \right)^2$	$\left(\frac{49u^2 - 4v^2}{14uv^2} \right)^2$	1
39	0,47	Rechne soweit als möglich: $\log_{0,1} 100$	$x = -2$	1
40	0,39	Rechne soweit als möglich: $3 \log_3 \frac{1}{3}$	$x = -3$	1

[illegible]

Güte der Anpassung an die Normalverteilung

Klassen- grenze	z	Fläche %	f_e	f_b	$(f_b - f_e)$	$(f_b - f_e)^2$	$\frac{(f_b - f_e)^2}{f_e}$
16,20	+0,841	20	12,4	12	-0,4	0,16	0,01
14,51	+0,253	20	12,4	13	+0,6	0,36	0,03
13,05	-0,253	20	12,4	11	-1,4	1,96	0,16
11,36	-0,841	20	12,4	12	-0,4	0,16	0,01
		20	12,4	14	+1,6	2,56	0,21

M = 13,78; S = 2,88; N = 62

$\chi^2 = 0,42$

Gruppe A knowledge / Frontalunterricht

Klassen- grenze	z	Fläche %	f_e	f_b	$(f_b - f_e)$	$(f_b - f_e)^2$	$\frac{(f_b - f_e)^2}{f_e}$
17,27	+0,841	20	12,4	12	-0,4	0,16	0,01
15,45	+0,253	20	12,4	17	+4,6	21,16	1,71
13,88	-0,253	20	12,4	14	+1,6	2,56	0,21
12,26	-0,841	20	12,4	6	-6,4	40,96	3,30
		20	12,4	13	+0,6	0,36	0,03

M = 14,67; S = 3,11; N = 62

$\chi^2 = 5,26$

Gruppe A knowledge / Progr. Unterricht

Güte der Anpassung an die Normalverteilung

Klassen- grenze	z	Fläche %	f_e	f_b	$(f_b - f_e)$	$(f_b - f_e)^2$	$\frac{(f_b - f_e)^2}{f_e}$
5,33	+0,841	20	12,4	7	-5,4	29,16	2,35
4,08	+0,253	20	12,4	13	+0,6	0,36	0,03
3,00	-0,253	20	12,4	11	-1,4	1,96	0,16
1,73	-0,841	20	12,4	20	+7,6	57,76	4,66
		20	12,4	11	-1,4	1,96	0,16

$$M = 3,54; S = 2,13; N = 62$$

$$\chi^2 = 7,36$$

Gruppe A comprehension / Frontalunterricht

Klassen- grenze	z	Fläche %	f_e	f_b	$(f_b - f_e)$	$(f_b - f_e)^2$	$\frac{(f_b - f_e)^2}{f_e}$
6,17	+0,841	20	12,4	14	+1,6	2,56	0,21
4,77	+0,253	20	12,4	6	-6,4	40,96	3,30
3,57	-0,253	20	12,4	10	-7,4	5,76	0,46
2,17	-0,841	20	12,4	19	+6,6	43,56	3,51
		20	12,4	13	+0,6	0,36	0,03

$$M = 4,17; S = 2,38; N = 62$$

$$\chi^2 = 7,51$$

Gruppe A comprehension / Progr. Unterricht

Güte der Anpassung an die Normalverteilung

Klassen- grenze	z	Fläche %	f_e	f_b	$(f_b - f_e)$	$(f_b - f_e)^2$	$\frac{(f_b - f_e)^2}{f_e}$
17,22	+0,841	20	11,4	14	+2,6	6,76	0,59
15,02	+0,253	20	11,4	11	-0,4	0,16	0,01
13,14	-0,253	20	11,4	10	-1,4	1,96	0,17
10,94	-0,841	20	11,4	9	-2,4	5,76	0,51
		20	11,4	13	+1,6	2,56	0,22

M = 14,08; S = 3,73; N = 57

$\chi^2 = 1,50$

Gruppe B knowledge / Frontalunterricht

Klassen- grenze	z	Fläche %	f_e	f_b	$(f_b - f_e)$	$(f_b - f_e)^2$	$\frac{(f_b - f_e)^2}{f_e}$
17,54	+0,841	20	11,4	10	-1,4	1,96	0,17
15,94	+0,253	20	11,4	15	+3,6	12,96	1,14
14,56	-0,253	20	11,4	12	+0,6	0,36	0,03
12,96	-0,841	20	11,4	8	-3,4	11,56	1,01
		20	11,4	12	+0,6	0,36	0,03

M = 15,25; S = 2,72; N = 57

$\chi^2 = 2,38$

Gruppe B knowledge / Progr. Unterricht

Güte der Anpassung an die Normalverteilung

Klassen- grenze	z	Fläche %	f _e	f _b	(f _b -f _e)	(f _b -f _e) ²	$\frac{(f_b - f_e)^2}{f_e}$
5,04	+0,841	20	11,4	8	-3,4	11,56	1,01
3,70	+0,253	20	11,4	14	+2,6	6,67	0,59
2,54	-0,253	20	11,4	9	-2,4	5,76	0,51
1,20	-0,841	20	11,4	14	+2,6	6,67	0,59
		20	11,4	12	+0,6	0,36	0,03

$$M = 3,12; S = 2,29; N = 57$$

$$\chi^2 = 2,73$$

Gruppe B comprehension / Frontalunterricht

Klassen- grenze	z	Fläche %	f _e	f _b	(f _b -f _e)	(f _b -f _e) ²	$\frac{(f_b - f_e)^2}{f_e}$
6,83	+0,841	20	11,4	11	-0,4	0,16	0,01
4,89	+0,253	20	11,4	8	-3,4	11,56	1,01
3,23	-0,253	20	11,4	12	+0,6	0,36	0,03
1,29	-0,841	20	11,4	14	+2,6	6,76	0,59
		20	11,4	12	+0,6	0,36	0,03

$$M = 4,06; S = 3,29; N = 57$$

$$\chi^2 = 1,67$$

Gruppe B comprehension / Progr. Unterricht

Güte der Anpassung an die Normalverteilung

Klassen- grenze	z	Fläche %	f_e	f_b	$(f_b - f_e)$	$(f_b - f_e)^2$	$\frac{(f_b - f_e)^2}{f_e}$
7,58	+0,841	20	12,4	12	-0,4	0,16	0,01
6,51	+0,253	20	12,4	11	-1,4	1,96	0,16
5,59	-0,253	20	12,4	16	+3,6	12,96	1,05
4,52	-0,841	20	12,4	9	-3,4	11,56	0,93
		20	12,4	14	+1,6	2,56	0,21

M = 6,05; S = 1,83; N = 62

$\chi^2 = 2,36$

Gruppe A Vorwissen

Klassen- grenze	z	Fläche %	f_e	f_b	$(f_b - f_e)$	$(f_b - f_e)^2$	$\frac{(f_b - f_e)^2}{f_e}$
7,80	+0,841	20	11,4	15	+3,6	12,96	1,14
6,60	+0,253	20	11,4	5	-6,4	40,96	3,59
5,58	-0,253	20	11,4	13	+1,6	2,56	0,22
4,38	-0,841	20	11,4	10	-1,4	1,96	0,17
		20	11,4	14	+2,6	6,76	0,59

M = 6,09; S = 2,03; N = 57

$\chi^2 = 5,71$

Gruppe B Vorwissen

Güte der Anpassung an die Normalverteilung

Klassen- grenze	z	Fläche %	f_e	f_b	$(f_b - f_e)$	$(f_b - f_e)^2$	$\frac{(f_b - f_e)^2}{f_e}$
124,29	+0,841	20	12,4	15	+2,6	6,76	0,55
119,25	+0,253	20	12,4	7	-5,4	29,16	2,35
114,90	-0,253	20	12,4	12	-0,4	0,16	0,01
109,84	-0,841	20	12,4	17	+4,6	21,16	1,70
		20	12,4	11	-1,4	1,96	0,16

M = 117,08; S = 8,58; N = 62

$\chi^2 = 4,77$

Gruppe A Gesamtintelligenz

Klassen- grenze	z	Fläche %	f_e	f_b	$(f_b - f_e)$	$(f_b - f_e)^2$	$\frac{(f_b - f_e)^2}{f_e}$
114,28	+0,841	20	12,4	12	-0,4	0,16	0,13
109,43	+0,253	20	12,4	10	-2,4	5,76	0,46
105,25	-0,253	20	12,4	19	+6,6	43,56	3,51
100,40	-0,841	20	12,4	6	-6,4	40,96	3,30
		20	12,4	15	+2,6	6,76	0,55

M = 107,34; S = 8,25; N = 62

$\chi^2 = 7,95$

Gruppe A mathematische Intelligenz

Güte der Anpassung an die Normalverteilung

Klassen- grenze	z	Fläche %	f_e	f_b	$(f_b - f_e)$	$(f_b - f_e)^2$	$\frac{(f_b - f_e)^2}{f_e}$
123,54	+0,841	20	11,4	12	+0,6	0,36	0,03
118,17	+0,253	20	11,4	9	-2,4	5,76	0,51
113,55	-0,253	20	11,4	12	+0,6	0,36	0,03
108,18	-0,841	20	11,4	11	-0,4	0,16	0,01
		20	11,4	13	+1,6	2,56	0,22

M E 115,86; S = 9,13; N = 57

$\chi^2 = 0,80$

Gruppe B Gesamtintelligenz

Klassen- grenze	z	Fläche %	f_e	f_b	$(f_b - f_e)$	$(f_b - f_e)^2$	$\frac{(f_b - f_e)^2}{f_e}$
112,70	+0,841	20	11,4	9	-2,4	5,76	0,50
108,06	+0,253	20	11,4	12	+0,6	0,36	0,03
104,07	-0,253	20	11,4	12	+0,6	0,36	0,03
99,43	-0,841	20	11,4	11	-0,4	0,16	0,01
		20	11,4	13	+1,6	2,56	0,24

M = 106,07; S = 7,89; N = 57

$\chi^2 = 0,81$

Gruppe B mathematische Intelligenz

Güte der Anpassung an die Normalverteilung

Klassen- grenze	z	Fläche %	f_e	f_b	$(f_b - f_e)$	$(f_b - f_e)^2$	$\frac{(f_b - f_e)^2}{f_e}$
23,37	+0,841	20	12,4	11	-1,4	1,96	0,16
20,48	+0,253	20	12,4	14	+1,6	2,56	0,21
18,00	-0,253	20	12,4	14	+1,6	2,56	0,21
15,11	-0,841	20	12,4	7	-5,4	29,16	2,35
		20	12,4	16	+3,6	12,96	1,05

M = 19,24; S = 4,91; N = 62

$\chi^2 = 3,98$

Gruppe A Einstellung zum Lehrer

Klassen- grenze	z	Fläche %	f_e	f_b	$(f_b - f_e)$	$(f_b - f_e)^2$	$\frac{(f_b - f_e)^2}{f_e}$
59,33	+0,841	20	12,4	16	+3,6	12,96	1,05
52,73	+0,253	20	12,4	9	-3,4	11,56	0,93
47,05	-0,253	20	12,4	14	+1,6	2,56	0,21
40,45	-0,841	20	12,4	10	-2,4	5,76	0,46
		20	12,4	13	+0,6	0,36	0,03

M = 49,89; S = 11,22; N = 62

$\chi^2 = 2,68$

Gruppe A Einstellung zur Schule

Güte der Anpassung an die Normalverteilung

Klassen- grenze	z	Fläche %	f _e	f _b	f _b -f _e	(f _b -f _e) ²	(f _b -f _e) ² f _e
10,94	+0,841	20	12,4	16	+3,6	12,96	1,05
9,81	+0,253	20	12,4	13	+0,6	0,36	0,03
8,83	-0,253	20	12,4	14	+1,6	2,56	0,21
7,70	-0,841	20	12,4	8	-4,4	19,36	1,56
		20	12,4	11	-1,4	1,96	0,16

$$M = 9,32; S = 1,93; N = 62$$

$$\chi^2 = 3,01$$

Gruppe A Arbeitshaltung (sozial)

Klassen- grenze	z	Fläche %	f_e	f_b	$(f_b - f_e)$	$(f_b - f_e)^2$	$\frac{(f_b - f_e)^2}{f_e}$
11,32	+0,841	20	12,4	10	-2,4	5,76	0,46
9,97	+0,253	20	12,4	16	+3,6	12,96	1,05
8,81	-0,253	20	12,4	16	+3,6	12,96	1,05
7,46	-0,841	20	12,4	9	-3,4	11,56	0,93
		20	12,4	11	-1,4	1,96	0,16

$$M = 9,39; S = 2,30; N = 62$$

$$\chi^2 = 3,65$$

Gruppe A Arbeitshaltung (konzentrativ)

Güte der Anpassung an die Normalverteilung

Klassen- grenze	z	Fläche %	f_e	f_b	$(f_b - f_e)$	$(f_b - f_e)^2$	$\frac{(f_b - f_e)^2}{f_e}$
23,91	+0,841	20	11,4	11	-0,14	0,16	0,01
20,54	+0,253	20	11,4	6	-5,14	29,16	2,56
17,64	-0,253	20	11,4	14	+2,6	6,76	0,59
14,27	-0,841	20	11,4	16	+4,6	21,16	1,89
		20	11,4	10	-1,4	1,96	0,17

M = 19,09; S = 5,73; N = 57

$\chi^2 = 5,19$

Gruppe B Einstellung zum Lehrer

Klassen- grenze	z	Fläche %	f_e	f_b	$(f_b - f_e)$	$(f_b - f_e)^2$	$\frac{(f_b - f_e)^2}{f_e}$
62,04	+0,841	20	11,4	9	-2,4	5,76	0,51
53,75	+0,253	20	11,4	14	+2,6	6,76	0,59
46,61	-0,253	20	11,4	8	-3,4	11,56	1,01
38,32	-0,841	20	11,4	14	+2,6	6,76	0,59
		20	11,4	12	+0,6	0,36	0,03

M = 50,18; S = 14,10; N = 57

$\chi^2 = 2,73$

Gruppe B Einstellung zur Schule

Güte der Anpassung an die Normalverteilung

Klassen- grenze	z	Fläche %	f_e	f_b	$(f_b - f_e)$	$(f_b - f_e)^2$	$\frac{(f_b - f_e)^2}{f_e}$
-11,81	+0,841	20	11,4	16	+4,6	21,16	1,86
-10,68	+0,253	20	11,4	8	-3,4	11,56	1,01
-9,70	-0,253	20	11,4	14	+2,6	6,76	0,59
-8,57	-0,841	20	11,4	10	-1,4	1,96	0,17
		20	11,4	9	-2,4	5,76	0,51

M = 10,19; S = 1,93; N = 57

$\chi^2 = 4,14$

Gruppe B Arbeitshaltung (sozial)

Klassen- grenze	z	Fläche %	f_e	f_b	$(f_b - f_e)$	$(f_b - f_e)^2$	$\frac{(f_b - f_e)^2}{f_e}$
-11,32	+0,841	20	11,4	13	+1,6	2,56	0,22
-9,71	+0,253	20	11,4	12	+0,6	0,36	0,03
-8,33	-0,253	20	11,4	6	-5,4	29,16	2,56
-6,72	-0,841	20	11,4	14	+2,6	6,76	0,59
		20	11,4	12	+0,6	0,36	0,03

M = 9,02; S = 2,73; N = 57

$\chi^2 = 3,43$

Gruppe B Arbeitshaltung (konzentrativ)

Güte der Anpassung an die Normalverteilung

Klassen- grenze	z	Fläche %	f_e	f_b	$(f_b - f_e)$	$(f_b - f_e)^2$	$\frac{(f_b - f_e)^2}{f_e}$
14,96	+0,841	20	12,4	11	-1,4	1,96	0,16
11,99	+0,253	20	12,4	12	-0,4	0,16	0,01
9,44	-0,253	20	12,4	14	+1,6	2,56	0,20
6,47	-0,841	20	12,4	15	+2,6	6,76	0,55
		20	12,4	10	-2,4	5,76	0,46

M = 10,72; S = 5,05; N = 62

$\chi^2 = 1,38$

Gruppe A positive Verstärkung

Klassen- grenze	z	Fläche %	f_e	f_b	$(f_b - f_e)$	$(f_b - f_e)^2$	$\frac{(f_b - f_e)^2}{f_e}$
23,19	+0,841	20	11,4	8	-3,4	11,56	1,01
17,54	+0,253	20	11,4	13	+1,6	2,56	0,22
12,74	-0,253	20	11,4	12	+0,6	0,36	0,03
7,15	-0,841	20	11,4	13	+1,6	2,56	0,22
		20	11,4	11	-0,4	0,16	0,01

M = 15,14; S = 9,50; N = 57

$\chi^2 = 1,49$

Gruppe B positive Verstärkung

Güte der Anpassung an die Normalverteilung

Klassen- grenze	z	Fläche %	f_e	f_b	$(f_b - f_e)$	$(f_b - f_e)^2$	$\frac{(f_b - f_e)^2}{f_e}$
20,96	$\pm 0,841$	20	12,4	8	-4,4	19,36	1,56
16,73	$\pm 0,253$	20	12,4	11	-1,4	1,96	0,16
13,33	$-0,253$	20	12,4	19	+6,6	43,56	3,51
9,37	$-0,841$	20	12,4	13	+0,6	0,36	0,03
		20	12,4	11	-1,4	1,96	0,16

$$M = 15,03; S = 6,73; N = 62$$

$$\chi^2 = 5,42$$

Gruppe A stoffbezogene Schüleraktivität

Klassen- grenze	z	Fläche %	f_e	f_b	$(f_b - f_e)$	$(f_b - f_e)^2$	$\frac{(f_b - f_e)^2}{f_e}$
30,35	$\pm 0,841$	20	11,4	9	-2,4	5,76	0,51
23,26	$\pm 0,253$	20	11,4	12	+0,6	0,36	0,03
17,16	$-0,253$	20	11,4	8	-3,4	11,56	1,01
10,07	$-0,841$	20	11,4	17	+5,6	31,36	2,75
		20	11,4	11	-0,4	0,16	0,01

$$M = 20,21; S = 12,06; N = 57$$

$$\chi^2 = 4,31$$

Gruppe B stoffbezogene Schüleraktivität

Güte der Anpassung an die Normalverteilung

Klassen- grenze	z	Fläche %	f_e	f_b	$(f_b - f_e)$	$(f_b - f_e)^2$	$\frac{(f_b - f_e)^2}{f_e}$
7,67	+0,841	20	12,4	15	+2,6	6,76	0,55
6,65	+0,253	20	12,4	14	+1,6	2,56	0,21
5,77	-0,253	20	12,4	10	-2,4	5,76	0,46
4,75	-0,841	20	12,4	14	+1,6	2,56	0,21
		20	12,4	9	-3,4	11,56	0,93

M = 6,21; S = 1,74; N = 62

$\chi^2 = 2,36$

Gruppe A Schulängstlichkeit

Klassen- grenze	z	Fläche %	f_e	f_b	$(f_b - f_e)$	$(f_b - f_e)^2$	$\frac{(f_b - f_e)^2}{f_e}$
7,63	+0,841	20	11,4	16	+4,6	21,16	1,86
6,62	+0,253	20	11,4	10	-1,4	1,96	0,17
5,74	-0,253	20	11,4	11	-0,4	0,16	0,01
4,73	-0,841	20	11,4	10	-1,4	1,96	0,17
		20	11,4	10	-1,4	1,96	0,17

M = 6,18; S = 1,722; N = 57

$\chi^2 = 2,30$

Gruppe B Schulängstlichkeit

Güte der Anpassung an die Normalverteilung

Klassen- grenze	z	Fläche %	f_e	f_b	$(f_b - f_e)$	$(f_b - f_e)^2$	$\frac{(f_b - f_e)^2}{f_e}$
0,7968	+0,841	20	4	6	+2	4	1,00
0,7445	+0,253	20	4	1	-3	9	2,25
0,6995	-0,253	20	4	5	+1	1	0,25
0,6472	-0,841	20	4	5	+1	1	0,25
		20	4	3	-1	1	0,25

M = 0,7220; S = 0,089; N = 20

$\chi^2 = 4,00$

P-Werte logarithmieren knowledge

Klassen- grenze	z	Fläche %	f_e	f_b	$(f_b - f_e)$	$(f_b - f_e)^2$	$\frac{(f_b - f_e)^2}{f_e}$
0,8142	+0,841	20	4	6	+2	4	1,00
0,7501	+0,253	20	4	1	-3	9	2,25
0,6949	-0,253	20	4	3	-1	1	0,25
0,6306	-0,841	20	4	5	+1	1	0,25
		20	4	5	+1	1	0,25

M = 0,7225; S = 0,109; N = 20

$\chi^2 = 4,00$

P-Werte Wurzelrechnen knowledge

Verteilung der Schwierigkeitsindices der Tests zur
Erfassung der Lernzielkategorie "comprehension"

Klassengrenze Wu/comprehension	fe	fb	Klassengrenze log/comprehension	fe	fb
0,414	1	1	0,457	1	1
	1	3	0,403	1	1
0,371	1	0		1	2
0,342	1	0	0,365	1	1
0,319	1	1	0,335	1	0
0,297	1	0	0,308	1	1
0,278	1	1	0,283	1	0
0,259	1	0	0,259	1	2
0,241	1	3	0,237	1	0
0,222	1	0	0,213	1	2
0,203	1	1	0,189	1	0
0,182	1	1	0,162	1	0
0,159	1	1	0,132	1	3
0,130	1	3	0,096	1	2
0,0901	1	0	0,045	1	0

N = 15

Z = 6

P $\approx$ 0,2784

N = 15

Z = 6

P $\approx$ 0,2784

Lebenslauf

Am 15. Juni 1942 wurde ich als zweites Kind des Volksschullehrers Gotthilf Walter und seiner Ehefrau Anne-liese, geb. Baur, in Tübingen geboren. Nach vierjähri-ger Grundschule wurde ich in das Keplergymnasium in Tübingen eingeschult. Dort legte ich im Jahr 1962 die Reifeprüfung ab. In der Zeit vom April 1962 bis Septem-ber 1963 absolvierte ich meinen Grundwehrdienst beim Panzerbataillon 304 in Münsingen. Anschließend inskri-bierte ich mich an der Universität Tübingen für das Hauptfachstudium Psychologie. Dort legte ich Ende des Wintersemesters 1965/66 die Vordiplomprüfung ab. Im Sommersemester 1966 inskribierte ich mich an der Uni-versität Erlangen-Nürnberg, im Sommersemester 1967 be-endete ich mein Studium mit dem Erwerb des Titels eines Diplompsychologen. Anschließend übernahm ich an der Päd-agogischen Hochschule Bayreuth die Verwaltung der Dienst-geschäfte eines wissenschaftlichen Assistenten. Seit Oktober 1969 bin ich Verwalter der Dienstgeschäfte eines wissenschaftlichen Assistenten an der Pädagogischen Hoch-schule München.

München, den 21. April 1971

Hellmuth Walter

Offsetdruck-Fotodruck

J. Hogl, 8520 Erlangen Hauptstr. 109, (Nähe Martin-Luther-Platz) Tel. 22990